

AI基础：论文写作工具

九老师 机器学习初学者 昨天

导语

现在工业界发paper越来越普遍了，AI从业人员要不想被淘汰，得学习。工欲善其事必先利其器，今天就讲讲用到的工具。

本文作者：九老师

出处：<https://zhuanlan.zhihu.com/p/100931497>

编辑：黄海广



目前我在编写AI基础系列，**目前已经发布：**

AI 基础：简易数学入门

AI 基础：Python开发环境设置和小技巧

AI 基础：Python 简易入门

AI 基础：Numpy 简易入门

AI 基础：Pandas 简易入门

AI 基础：Scipy(科学计算库) 简易入门

AI基础：数据可视化简易入门（matplotlib和seaborn）

AI基础：机器学习库Scikit-learn的使用

AI基础：机器学习简易入门

AI基础：机器学习的损失函数

AI基础：机器学习和深度学习的练习数据

AI基础：特征工程-类别特征

AI基础：特征工程-数字特征处理

AI基础：特征工程-文本特征处理

AI基础：词嵌入基础和Word2Vec

AI基础：图解Transformer

AI基础：一文看懂BERT

AI基础：入门人工智能必看的论文

AI基础：走进深度学习

AI基础：优化算法

AI基础：卷积神经网络

AI基础：经典卷积神经网络

AI基础：深度学习论文阅读路线（127篇经典论文下载）

AI基础：数据增强方法综述

后续持续更新

工具使用流程

先讲讲本文提到的工具使用的流程。

在**Overleaf**使用**Latex**进行论文编辑，随时可以转化成PDF看效果；写一句话时，塑料英语不知道如何下笔，谷歌翻译中翻英帮忙确立句子轮廓；粘贴到**Grammarly**中进行修改，过程中根据语法错误提示，避免低级错误，对那些常用词进行同义替换（常见词往往不精准）；如果遇到了一些词语不知道如何搭配，**Linggle**通过通配符匹配，给出文献中使用的频率和例句。

为了写作，必须要阅读大量的论文。谷歌在大多数都能轻松找到PDF，偶尔遇到收费的，这时候到**SCI-HUB**上碰碰运气；大量阅读过的论文，通过**Mendeley**来管理，它除了一个PDF阅读器的基本功能，按文献的维度（如作者，会议，年份）结构化和检索，方便的论文导入，多端同步；需要引用的时候，直接从**mendeley**里导出BibTex格式的片段；想要写公式的时候，自己用**latex**搭出公式来挺费劲的，在别人论文里见过类似的，**Mathpix**可以轻松把图片转化为**latex**源码，略作修改就用到自己文章里了。

下面逐一给出简要的介绍，和工具的地址。

在线的**latex**编辑和编译工具：**overleaf**

论文最终展现出来的就是一个PDF格式的文档。

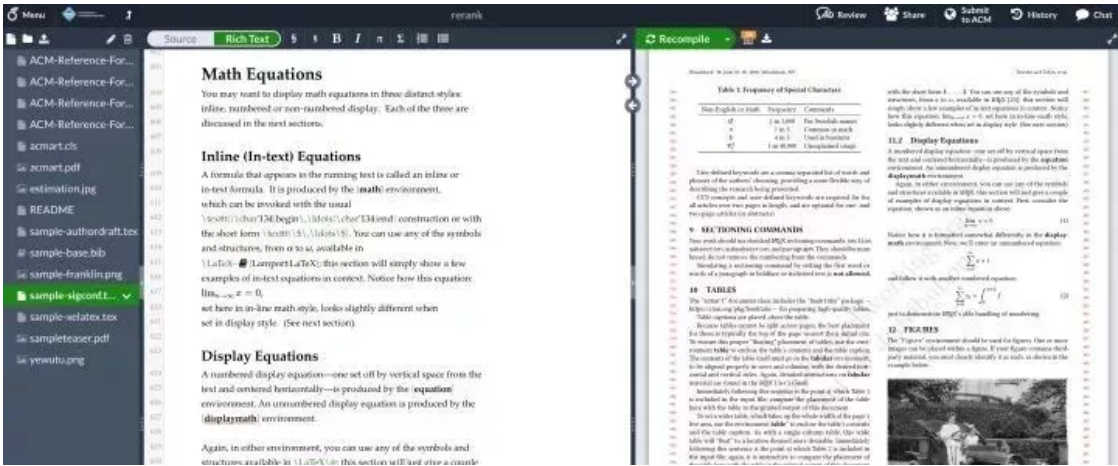
当然可以使用**word**，但光排版这件事情，就能耗费你一半的精力。

正确的答案是，使用**latex**，它是一个专业的排版工具，按照**latex**的语法进行写作，执行编译就能够得到PDF文件。它的语法包含了如何排版，虽然相比**word**上手要慢，但在排版这件事情上，入门级别的**latex**语法，你要达到精通**word**的水平。

latex如何使用呢？当然，要安装编译器，再安装编辑器，本地一通配置，偶尔会遇到些问题，凭着强大的谷歌搜索，倒也不是什么难事。配置本地环境，不如直接使用在线编辑器。

www.overleaf.com

- 注册即用，免去本地**latex**环境安装的痛苦。
- 多人合作，共同编辑。
- 富文本编辑模式，比写**latex**源码舒服些。
- 随时可以完成在线编译，查看PDF。



按照overleaf的开始流程，有选择模板的过程，模板怎么选，还是要看投稿的期刊或者会议的要求。以KDD为例，在它的KDD 2019 Call for Research Papers页面上，给出了模板格式，看看能不能在overleaf上找到，即使没有，一会提供下载，自己上传到overleaf。

Submission Directions

KDD is a dual track conference hosting both a Research track and an Applied Data Science track. Due to the large number of submissions, papers submitted to the Research track will not be considered for publication in the Applied Data Science track and vice versa. Authors are encouraged to read the track descriptions carefully and to choose an appropriate track for their submissions. Submissions are limited to a total of nine (9) pages, including all content and references, and must be in PDF format and formatted according to the new Standard ACM Conference Proceedings Template. For LaTeX users: unzip acmart.zip, make, and use sample-sigconf.tex as a template;

Additional information about formatting and style files is available online at: <https://www.acm.org/publications/proceedings-template>.

Papers that do not meet the formatting requirements will be rejected without review. In addition, authors can provide an optional two (2) page supplement at the end of their submitted paper (it needs to be in the same PDF file and start at page 10) focused on reproducibility. This supplement can only be used to include (i) information necessary for reproducing the experimental results, insights, or conclusions reported in the paper (e.g., various algorithmic and model parameters and configurations, hyper-parameter search spaces, details related to dataset filtering and train/test splits, software versions, detailed hardware configuration, etc.), and (ii) any pseudo-code, or proofs that due to space limitations, could not be included in the main nine-page manuscript, but that help in reproducibility (see reproducibility policy below for more details).

The Research track follows a double-blind review process. Submitted papers must not include author names and affiliations and they must be written in a way so that they do not break the double-blind reviewing process. Papers that violate the double-blind review requirements will be desk rejected.

Because of the double-blind review process, non-anonymous papers that have been issued as technical reports or similar cannot be considered for KDD 2019. An exception to this rule applies to arXiv papers that were published in arXiv at least a month prior to KDD's submission deadline (January 3, 2019). Authors can submit these arXiv papers to KDD provided that the submitted paper title and abstract is different from the one appearing on arXiv. Papers that appear on arXiv from January 3, 2019

走过这一步，已经可以编译出模板PDF了，可以照猫画虎地写起来了。

谷歌翻译

开始写作前，要纠结一个问题，是先写中文再翻译，还是直接写英文？我觉得，怎么快怎么来，避免拖延，快速开始。英语不过关，翻译软件来凑。

有时候不知道怎么开始一个句型，先中文拿到英文句型的轮廓，再靠着自己的塑料英文加上更多的辅助工具完善，不失为一个好的选择。

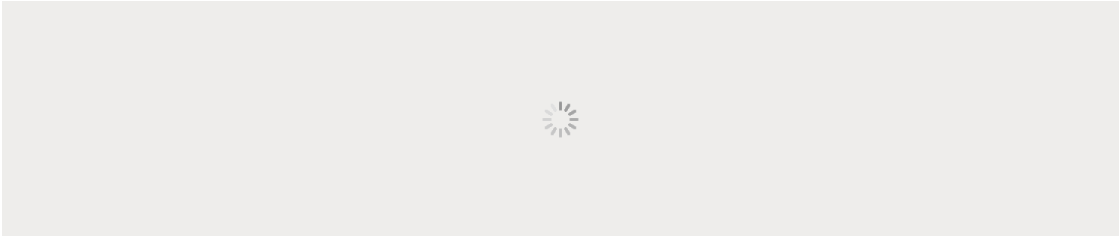
translate.google.cn

一个秘传用法：中翻英，改英文，英翻中，改中文。。。反复迭代

grammarly: 在线的语法纠错神器

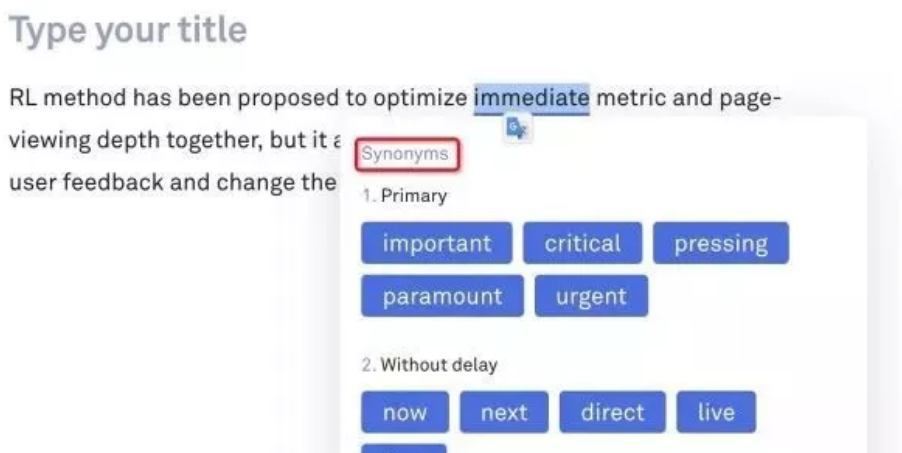
<https://app.grammarly.com/>

在这编辑文章的一句或一段话，语法出错了会有提示，低级的语法错误都能够避免。



语法纠错

除了语法纠错之外，还有同意替换功能，我的塑料英语能想到的词汇都太过常见，不够精准（逼格不足），选中词就可以同义替换了。



同意替换

linggle:解决词语搭配问题

上面的grammarly可以语句的粗修，而它并不解决词汇搭配的问题，哪些短语自然，意思更精准，linggle就能完成词汇的精修。

通过短语加通配符，它能够匹配到合适的短语，填补空缺，并给出词频和例句。

Linggle:

Engineslinggle.com

除了下图使用了下划线匹配任意词，还有更加复杂的通配符，在首页就有教程，简单易用。

Phrase	Percent	Count	Example
demonstrate the use of	9.8 %	51,420	Show
demonstrate the importance of	7.2 %	37,806	Show
demonstrate the effectiveness of	6.7 %	35,220	Hide

• To demonstrate the effectiveness of good technique I invited someone-anyone to jump in the water and grab me from @@@ and was confident that I could handle anything ; after all , everyone is only a head tall in the water .
 • This method provides a way to evaluate and determine the medical cost savings resulting from social work interventions and to demonstrate the effectiveness of social work services in a managed care setting .
 • BIOREMEDIATION TECHNOLOGY Projects demonstrate the effectiveness of composting as a bioremediation tool .
 • Throughout the 1960s and 1970s , math contests , district- and state-level science fairs , poetry contests , and a host of other events grew in popularity in an effort to provide various means of public recognition that would motivate students to excel and would demonstrate the effectiveness of these educational activities .
 • Section : Search and Discovery To demonstrate the effectiveness of neutron diffraction in biology , crystallographers bring neutrons to bear on an important industrial enzyme .
 • To demonstrate the effectiveness of his method , he embalmed an unclaimed body at the Eclectic Medical College of New York , injecting it with several gallons of his secret fluid via the carotid artery .
 • Two numerical examples are worked out to demonstrate the effectiveness of the proposed design technique .
 • The freebies are used to demonstrate the effectiveness of their site design -- its accessibility , its organizational clarity , its straightforward navigating cues , and its downloading speed .
 • A trial conducted in 1947-48 by Britain 's Medical Research Council , however , helped demonstrate the effectiveness of the antibiotic streptomycin .
 • In contrast to statements from the state 's education commissioner , which imply that music educators @@@ a good concert would do more to demonstrate the effectiveness of a competent music teacher than the mounds of paperwork generated by Race to the Top .

demonstrate the value of	6.6 %	34,896	Show
--------------------------	-------	--------	------

SCI-HUB：论文检索

谷歌都找不到PDF的时候，用sci-hub试试。

scihub最新可用地址

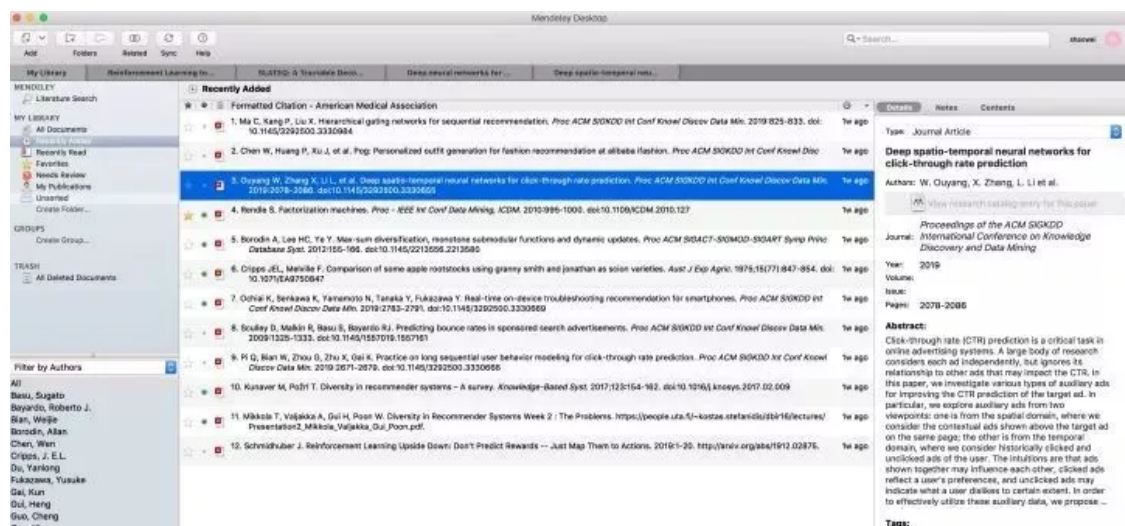
www.sci-hub.shop

论文阅读器：mendeley

写作前，一定要大量阅读相关工作，一个靠谱的论文管理和阅读工具非常重要。

<https://www.mendeley.com>

- 论文结构化和索引，可以自己打标签，方便检索。
- 论文链接粘贴复制导入。
- 多端同步。
- 论文引用，BibTex格式获取。



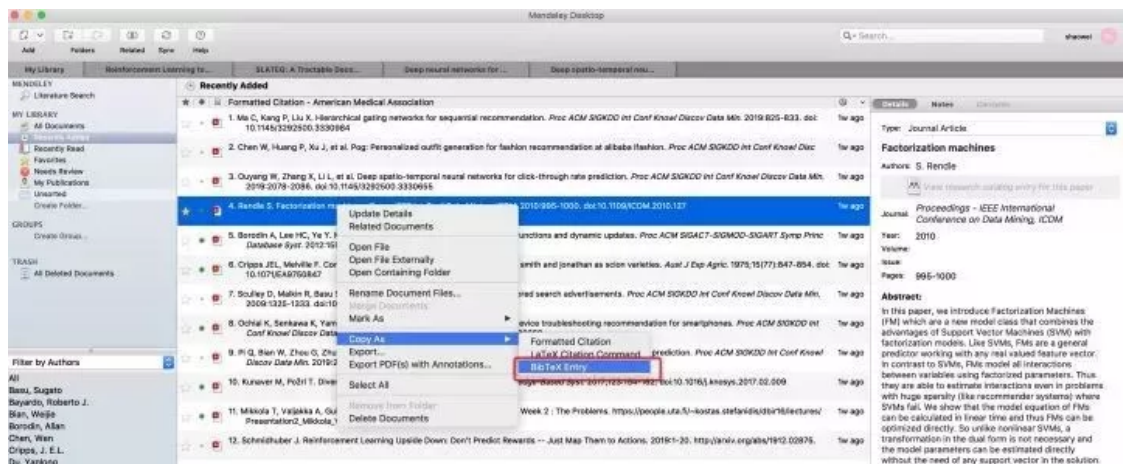
在论文中引用相关工作是不可避免的。latex在线编辑工具中，有一个bib后缀的文件，里面有类似下边的片段。

```
@Article{Abril07,
  author      = "Patricia S. Abril and Robert Plant",
  title       = "The patent holder's dilemma: Buy, sell, or troll?",
  journal     = "Communications of the ACM",
  volume      = "50",
  number      = "1",
  month       = jan,
  year        = "2007",
  pages       = "36--44",
  doi         = "10.1145/1188913.1188915",
  url         = "http://doi.acm.org/10.1145/1219092.1219093",
  note       = "",
}
```

在引用的时候，latex文档中使用\cite{Abril07}就可以引用了。

那这个片段是哪里来的？

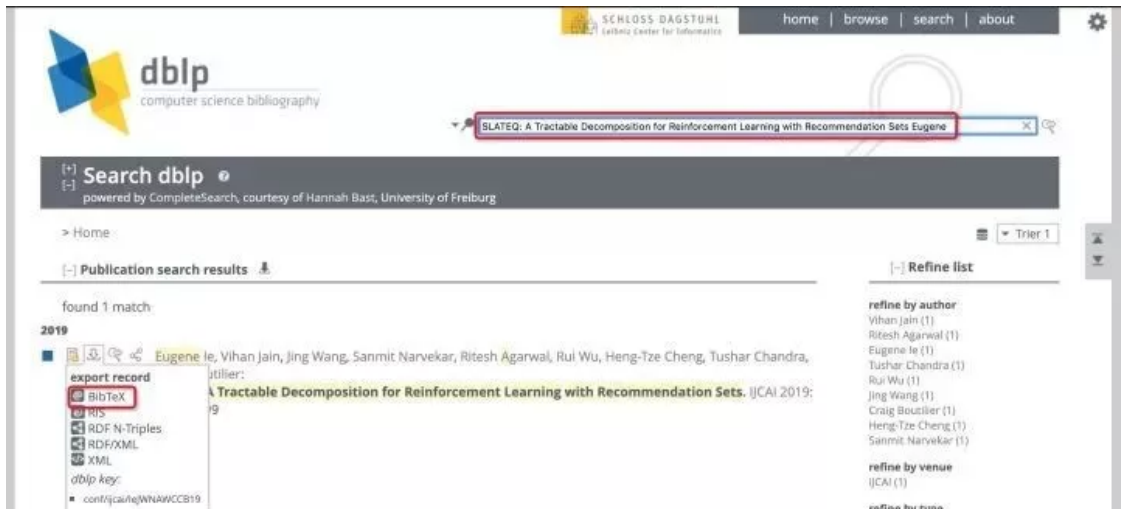
可以使用上面提到的**mendeley**，方便地获取BibTex引用片段。



或者，推荐一个网站：

<https://dblp.uni-trier.de/>

搜索论文题目，点击BibTex，就能看到那一段引用片段了。

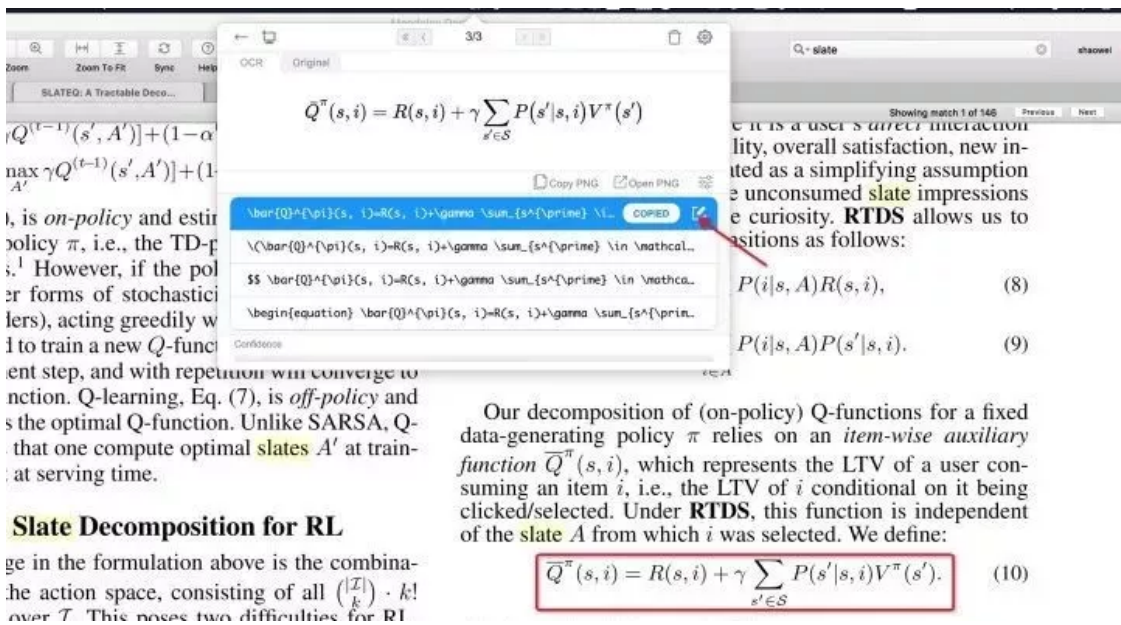


相似工具：**zotero**（见文章介绍）

mathpix：公式图片转latex

数学公式转latex，有时候自己写个公式还挺费劲的，现在轻松截图，公式在手。

[Snipmathpix.com](https://www.snipmathpix.com)



备注：mathpix已经收费，免费用户每个月只能截图50次。

解决方法：

1. 付费充值；
2. 用多个邮箱注册，每个邮箱可以用50次，正常用途一个月300-400次够了，用8个左右邮箱。

参考资料

[1] <https://zhuanlan.zhihu.com/p/100931497>



往期精彩回顾



- 2019年公众号文章精选
- 适合初学者入门人工智能的路线及资料下载
- 机器学习在线手册
- 深度学习在线手册
- AI基础下载（第一部分）

备注：加入本站微信群或者qq群，请回复“加群”

加入知识星球（4500+用户，ID：92416895），请回复“知识星球”

喜欢文章，点个在看👉

