

华东师范大学本科生毕业论文（设计）开题报告

论文题目	文档标签生成软件设计与实现	数据科学与工程学院		数据科学与大数据技术专业	
		学生姓名	蔡闻宇	学号	10174900160
一、选题的背景与意义					
选题背景： <p>文件管理一直是计算机操作系统中的重要组成部分。市面上主流操作系统均采用树形目录结构来管理检索和显示文件[1]。树形目录结构在有许多优点的同时，也有一定的缺陷，比如不适合处理逻辑划分模糊，重叠的文件。此时，标签式文件管理软件也开始进入大家视野。标签式文件管理通过标签来组织与检索文件[2]。</p> <p>而标签式文件管理有一个问题：标签人为录入较为繁琐，每个文件都需要用户手动添加所有的标签。调查了市面上较为流行的标签式文件管理软件，发现尽管部分软件通过优秀的UI设计来缓解这一问题，但是均未实质性解决现存的使用困难。</p> <p>同时，在调查国内外有关于标签式文件管理方法以及手动输入标签的相关研究[3][4][5][6][7]，发现对于本文提出的问题相关研究较少，这个问题具有应用价值也同时具有研究价值。</p>					
选题意义： <ol style="list-style-type: none">1. 针对当前“手动输入所有标签”的问题，采用机器学习的方法，设计自动化标签生成方案，有效管理用户文件。2. 基于该想法，利用传统统计学习技术，实现一款轻量级的文件标签式管理软件。					

二、研究的主要内容和预期目标

本次研究的主要内容：

1. 轻量级标签式文件管理软件的设计与实现。
2. 探索发现能够解决“手动添加标签”这一难点问题的方法，支持自动化生成有效标签。

本次研究的预期目标如下：

实现一个能够自动添加标签的轻量级文件管理软件。

完成：

1. 实现一个标签式轻量级文件管理软件。
2. 实现一种能自动生成有效标签的算法。
3. 确保生成的标签有效，使用户能更高效地维护与管理标签。

三、拟采用的研究方法、步骤

本次研究主要采用的方法是文献研究法和实验法。

1. 文献研究法：

详细调研国外对于标签文件管理的相关研究[3][4][5][6][7]，同时参考国内已有的标签式文件管理软件的设计方案[8][9]，最后得出最终的软件设计方案。

2. 实验法：

在充分调研的基础上，实现该标签式文件管理软件。

四、研究的总体安排与进度

2020.11-2020.12.7	确定题目
2020.12.8-2021.1.4	撰写开题报告
2021.1.5-2021.2.1	实验设计与准备
2021.2.2-2021.3.1	开展实验
2021.3.2-2021.3.20	撰写中期报告，论文初稿
2021.3.21-2021.4.10	完成论文
2021.4.11-2021.5.1	完成查重等论文完善工作，论文定稿
2021.5.2-2021.5	准备答辩

五、参考文献

- [1] Dinneen, J. D., & Julien, C. (2019). *The ubiquitous digital file: A review of file management research. Journal of the Association for Information Science and Technology*. doi:10.1002/asi.24222
- [2] Hsieh, Wen-Tai, and Wei-Shen Lai. "Tag organization methods and systems." U.S. Patent Application No. 11/641,699.
- [3] Albadri, Nehad, Richard Watson, and Stijn Dekeyser. "TreeTags: bringing tags to the hierarchical file system." *Proceedings of the Australasian Computer Science Week Multiconference*. 2016.
- [4] Adrian, Benjamin, Leo Sauermann, and Thomas Roth-Berghofer. "Contag: A semantic tag recommendation system." *Proceedings of I-Semantics 7* (2007): 297-304.
- [5] Bloehdorn, Stephan, et al. "Tagfs-tag semantics for hierarchical file systems." *Proceedings of the 6th International Conference on Knowledge Management (I-KNOW 06), Graz, Austria*. Vol. 8. 2006.
- [6] Oleksik, Gerard, et al. "Lightweight tagging expands information and activity management practices." *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2009.
- [7] Voit, Karl, Keith Andrews, and Wolfgang Slany. "Tagging might not be slower than filing in folders." *CHI'12 Extended Abstracts on Human Factors in Computing Systems*. 2012. 2063-2068.
- [8] 尹华国. 基于标签的内容管理系统设计 [D]. 四川: 西南交通大学, 2010. DOI:10.7666/d.y1956268.
- [9] 郭彩云, 王会进, 龙舜, 等. 面向中小机构的文件管理系统的设计与实现 [J]. 计算机应用与软件, 2016, 33 (9) : 72-75, 87. DOI:10.3969/j.issn.1000-386x.2016.09.017.

论文题目	文档标签生成软件设计与实现	数据科学与工程学院		数据科学与大数据技术专业	
		学生姓名	蔡闻宇	学号	10174900160

六、指导教师意见

蔡闻宇同学的论文立足于实际应用，基于已有知识，解决现实应用问题。因此选题合适。根据对已有工具的总结，该题目具有一定的挑战性。

前期该同学已经做了较为广泛的调研，参考了国内外学术研究论文，实际体验了现有的标签管理软件，对问题也具有了一定的认知。论文将以标签生成的自动化、有效化和通用化为目标，设计可用性高的标签生成软件系统。

综上所述，论文选题良好，问题定义较为清晰，对于本科生来说具有一定的挑战性。同意蔡闻宇同学开题。

签字：	张蓉	2021年	1月	5日
-----	----	-------	----	----

七、开题答辩小组意见

针对蔡闻宇同学的报告，小组有以下意见：1. 希望之后的工作，能在更加限定的范围内开展，原范围可能对于本科生过难。2. 在结合上一点的前提下，报告后调整的工作能保证科研深度。

综上，同意进一步开展工作。

小组成员签字：	原晓梅 王芳 2021年	1月	5日
---------	--------------	----	----

教务处编制