

题 目： 基于人工策略模型及机器学习算法的网页分类器模型的设计与过滤系统的实现

学院： 软件学院 专业： 软件工程 学生姓名： XXX 学号： XXXXXXXX

指导教师（签名）： _____ 提交日期： 2016 年 3 月 4 日

毕业设计（论文）基本内容和要求：

毕设内容：

学生通过实习项目中学习设计开发网页类型识别模型，其中包括有人工策略模型及机器学习模型。理论方面，学生了解了搜索引擎的原理和 workflow，以及设计模型所需要的知识技能，如所涉及到的机器学习算法。学习了项目开发的流程：从调研开始，对项目进行需求分析，预期收益效果分析，设计模型，开发模型，至项目完成后对项目的收益计算。学生在完成实习项目后，应多深入研究模型开发设计及发现模型其他潜在应用并提出方案，编写论文及报告。

毕设要求：

要求学生加以研究如何更好的实现模型开发，并在满足高准确的要求下提升召回率。

要求学生继续发现模型的应用，不仅在实习项目中提及的用于识别垃圾网页，删除网页库中垃圾网页，并过滤垃圾网页的作用。还可能存在多方面的应用。

要求学生具备独立开发模型的能力，熟悉了解整个模型开发流程，并且熟悉掌握

开发模型流程中所必备的技术和理论基础。

毕业设计（论文）重点研究的问题：

1. 研究网页分类器模型的设计与开发方案
2. 研究如何设计网页过滤系统
3. 研究如何设法提升模块效率，通过调研实验，优化现有模块
4. 发现现有技术（通过在已有模块下开发页面类型识别模型）在其他方面的应用，及分析可能的潜在收益
5. 研究其他机器学习算法对模型开发的应用效果

毕业设计（论文）应完成的工作：

1. 阅读相关文献，并且翻译相关资料（不少于一万字）。
2. 完成模型的设计与实现，并且书写相关的文档报告，如初期调研结果，需求分析，预期收益分析，模型设计方案，采用的机器学习算法介绍等文档。
3. 网页类型识别模型
4. 编写毕业论文，不少于两万字

参考资料推荐：

1. 搜索引擎：信息检索实践 [Search Engines Information Retrieval in Practice] [美] 克罗夫特 等 著
2. 统计学原理 李洁明、祁新娥
3. 神经网络与机器学习(原书第3版) (加)海金
4. 网络上的一些参考资料。
5. 图书馆和网络上的一些参考文献，比如论文等。

其他要说明的问题：