李天宝

turingmac@hotmail.com

基于知识库的海量异构数据集成系统的设计与实现

实习周报

上海骇咕赛信息科技有限公司

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **本科生实习&毕业设计（论文）工作周报** | | | | | | | | |
| 学生姓名 | 李天宝 | 联系电话 | 15704600640 | | Email | turingmac@hotmail.com | | |
| 实习基地  (或实验室) | 上海骇咕赛信息科技有限公司 | | 基地导师 | 丁盛豪 | | | 校内导师 | 王宏志 |
| 报告日期 | 2016-07-24 | | 实习周次 | 第1周 | | | | |
| 报告区间 | 2016-07-18 到 2016-07-24 | | | | | | | |
| 工作重点 | 配置公司要求的环境 | | | | | | | |
| 工作内容及进度 | 1. Hadoop、Hbase环境配置 2. Boost库配置 3. 其他环境配置 | | | | | | | |
| 遇到问题 | 1. Linux权限问题导致不能运行，以及Hadoop的端口被占用 2. 私钥配置失败导致Hadoop不能启动 3. Linux空间不足，Hadoop无法启动 | | | | | | | |
| 自我小结 | 本周主要进行了基本环境的配置，由于对Linux不是很熟悉，在配置Hadoop的过程中出现了不少的问题 | | | | | | | |
| 下周计划 | 熟悉公司情况和业务流程 | | | | | | | |
| 导师意见 | 签字： | | | | | | | |

**说明：**

1. 上述表格大小可以根据文字多少调整；
2. 每周要写工作周报并发送给校内导师，实习基地导师可根据需要发送；
3. 校内实习学生每2周要书面报呈周报给导师，当面指导并签字。

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **本科生实习&毕业设计（论文）工作周报** | | | | | | | | |
| 学生姓名 | 李天宝 | 联系电话 | 15704600640 | | Email | turingmac@hotmail.com | | |
| 实习基地  (或实验室) | 上海骇咕赛信息科技有限公司 | | 基地导师 | 丁盛豪 | | | 校内导师 | 王宏志 |
| 报告日期 | 2016-07-31 | | 实习周次 | 第2周 | | | | |
| 报告区间 | 2016-07-25 到 2016-07-31 | | | | | | | |
| 工作重点 | 熟悉公司情况和业务流程 | | | | | | | |
| 工作内容及进度 | 1. 了解公司的集群，阅读相关文档 2. 了解公司的代码规范 3. 配置并使用公司的Git服务器 4. 学习使用团队协作软件 | | | | | | | |
| 遇到问题 | 1. 公司的代码规范与平时写代码的习惯差别较大 2. 公司内提交代码的代码审核有着既定的流程，需要适应 | | | | | | | |
| 自我小结 | 本周主要对公司日常的工作流程进行了解，融入公司的开发环境 | | | | | | | |
| 下周计划 | 熟悉Hadoop集群和公司现有集群 | | | | | | | |
| 导师意见 | 签字： | | | | | | | |

**说明：**

1. 上述表格大小可以根据文字多少调整；
2. 每周要写工作周报并发送给校内导师，实习基地导师可根据需要发送；
3. 校内实习学生每2周要书面报呈周报给导师，当面指导并签字。

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **本科生实习&毕业设计（论文）工作周报** | | | | | | | | |
| 学生姓名 | 李天宝 | 联系电话 | 15704600640 | | Email | turingmac@hotmail.com | | |
| 实习基地  (或实验室) | 上海骇咕赛信息科技有限公司 | | 基地导师 | 丁盛豪 | | | 校内导师 | 王宏志 |
| 报告日期 | 2016-08-07 | | 实习周次 | 第3周 | | | | |
| 报告区间 | 2016-08-01 到 2016-08-07 | | | | | | | |
| 工作重点 | 熟悉Hadoop集群和公司现有集群 | | | | | | | |
| 工作内容及进度 | 1. 了解公司的集群，阅读相关文档 2. 了解Hadoop，学习Hadoop streaming开发 | | | | | | | |
| 遇到问题 | 1. 自己对python语言掌握的不够熟练 2. 不清楚如何使用集群，本机测试通过的map-reduce代码补鞥在公司的集群上运行 | | | | | | | |
| 自我小结 | 本周主要对学习了Hadoop相关的知识，并开发了相关的一些小demo，算是在分布式处理方面开了个头 | | | | | | | |
| 下周计划 | Pig语言的学习 | | | | | | | |
| 导师意见 | 签字： | | | | | | | |

**说明：**

1. 上述表格大小可以根据文字多少调整；
2. 每周要写工作周报并发送给校内导师，实习基地导师可根据需要发送；
3. 校内实习学生每2周要书面报呈周报给导师，当面指导并签字。

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **本科生实习&毕业设计（论文）工作周报** | | | | | | | | |
| 学生姓名 | 李天宝 | 联系电话 | 15704600640 | | Email | turingmac@hotmail.com | | |
| 实习基地  (或实验室) | 上海骇咕赛信息科技有限公司 | | 基地导师 | 丁盛豪 | | | 校内导师 | 王宏志 |
| 报告日期 | 2016-08-14 | | 实习周次 | 第4周 | | | | |
| 报告区间 | 2016-08-08 到 2016-08-14 | | | | | | | |
| 工作重点 | Pig语言的学习 | | | | | | | |
| 工作内容及进度 | 1. 学习Pig语言，阅读官方文档 2. 浏览《Pig编程指南》 3. 写demo进行测试 | | | | | | | |
| 遇到问题 | 1. 使用Jython设计Pig的UDF在实际应用中，与2.0以上版本的hadoop存在兼容性bug 2. Hadoop2.0以上版本不再支持C语言的Pig UDF编程 | | | | | | | |
| 自我小结 | 本周主要对学习了Pig相关的知识，作为一门Hadoop平台之上设计的语言，提供了很多附加的功能，与实习所要设计的平台有些类似，应仔细研究 | | | | | | | |
| 下周计划 | 研究Hadoop源码 | | | | | | | |
| 导师意见 | 签字： | | | | | | | |

**说明：**

1. 上述表格大小可以根据文字多少调整；
2. 每周要写工作周报并发送给校内导师，实习基地导师可根据需要发送；
3. 校内实习学生每2周要书面报呈周报给导师，当面指导并签字。

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **本科生实习&毕业设计（论文）工作周报** | | | | | | | | |
| 学生姓名 | 李天宝 | 联系电话 | 15704600640 | | Email | turingmac@hotmail.com | | |
| 实习基地  (或实验室) | 上海骇咕赛信息科技有限公司 | | 基地导师 | 丁盛豪 | | | 校内导师 | 王宏志 |
| 报告日期 | 2016-08-21 | | 实习周次 | 第5周 | | | | |
| 报告区间 | 2016-08-15 到 2016-08-21 | | | | | | | |
| 工作重点 | 研究Hadoop源码 | | | | | | | |
| 工作内容及进度 | 1. 阅读《Hadoop核心技术》部分章节 2. 使用Java完成map-reduce开发 3. 使用Java开发Pig UDF | | | | | | | |
| 遇到问题 | 1. 对于Java的接口不太了解 2. Java中与Pig中数据结构类型存在一定的差异 3. 对于Java的高级数据结构不熟悉 | | | | | | | |
| 自我小结 | 本周主要对学习了使用原生语言Java开发Hadoop相关程序，这是开发Hadoop程序以及公司相关平台上程序必要的 | | | | | | | |
| 下周计划 | 使用公司平台的功能开发demo | | | | | | | |
| 导师意见 | 签字： | | | | | | | |

**说明：**

1. 上述表格大小可以根据文字多少调整；
2. 每周要写工作周报并发送给校内导师，实习基地导师可根据需要发送；
3. 校内实习学生每2周要书面报呈周报给导师，当面指导并签字。

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **本科生实习&毕业设计（论文）工作周报** | | | | | | | | |
| 学生姓名 | 李天宝 | 联系电话 | 15704600640 | | Email | turingmac@hotmail.com | | |
| 实习基地  (或实验室) | 上海骇咕赛信息科技有限公司 | | 基地导师 | 丁盛豪 | | | 校内导师 | 王宏志 |
| 报告日期 | 2016-08-28 | | 实习周次 | 第6周 | | | | |
| 报告区间 | 2016-08-22 到 2016-08-28 | | | | | | | |
| 工作重点 | 使用公司平台的功能开发demo | | | | | | | |
| 工作内容及进度 | 1. 使用公司平台接口开发，熟悉功能，在此保密 | | | | | | | |
| 遇到问题 | 1. 对公司的接口不熟悉 | | | | | | | |
| 自我小结 | 本周主要对学习了使用公司的接口进行开发测试程序，了解整个平台的工作原理，熟悉功能，并在使用过程中寻找可添加的功能 | | | | | | | |
| 下周计划 | 分析系统，寻找开发点 | | | | | | | |
| 导师意见 | 签字： | | | | | | | |

**说明：**

1. 上述表格大小可以根据文字多少调整；
2. 每周要写工作周报并发送给校内导师，实习基地导师可根据需要发送；
3. 校内实习学生每2周要书面报呈周报给导师，当面指导并签字。

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **本科生实习&毕业设计（论文）工作周报** | | | | | | | | |
| 学生姓名 | 李天宝 | 联系电话 | 15704600640 | | Email | turingmac@hotmail.com | | |
| 实习基地  (或实验室) | 上海骇咕赛信息科技有限公司 | | 基地导师 | 丁盛豪 | | | 校内导师 | 王宏志 |
| 报告日期 | 2016-09-04 | | 实习周次 | 第7周 | | | | |
| 报告区间 | 2016-08-29 到 2016-09-04 | | | | | | | |
| 工作重点 | 分析系统，寻找开发点 | | | | | | | |
| 工作内容及进度 | 1. 阅读公司平台设计文档，了解架构 | | | | | | | |
| 遇到问题 | 1. 系统过于复杂 | | | | | | | |
| 自我小结 | 本周主要对公司的系统进行一定的了解 | | | | | | | |
| 下周计划 | 分析系统，寻找开发点 | | | | | | | |
| 导师意见 | 签字： | | | | | | | |

**说明：**

1. 上述表格大小可以根据文字多少调整；
2. 每周要写工作周报并发送给校内导师，实习基地导师可根据需要发送；
3. 校内实习学生每2周要书面报呈周报给导师，当面指导并签字。

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **本科生实习&毕业设计（论文）工作周报** | | | | | | | | |
| 学生姓名 | 李天宝 | 联系电话 | 15704600640 | | Email | turingmac@hotmail.com | | |
| 实习基地  (或实验室) | 上海骇咕赛信息科技有限公司 | | 基地导师 | 丁盛豪 | | | 校内导师 | 王宏志 |
| 报告日期 | 2016-09-11 | | 实习周次 | 第8周 | | | | |
| 报告区间 | 2016-09-05 到 2016-09-11 | | | | | | | |
| 工作重点 | 分析系统，寻找开发点 | | | | | | | |
| 工作内容及进度 | 1. 阅读公司平台部分源码 2. 学习部分接口API 3. 了解数据集成相关内容 | | | | | | | |
| 遇到问题 | 1. 想做的数据集成在现有的平台上没有先例 2. 对于数据集成方面没有经验 | | | | | | | |
| 自我小结 | 本周主要基于对公司平台的了解，提出一些自己想做的内容，目前的关注点是数据集成，之后将验证这个想法的可行性 | | | | | | | |
| 下周计划 | 选择研究方向 | | | | | | | |
| 导师意见 | 签字： | | | | | | | |

**说明：**

1. 上述表格大小可以根据文字多少调整；
2. 每周要写工作周报并发送给校内导师，实习基地导师可根据需要发送；
3. 校内实习学生每2周要书面报呈周报给导师，当面指导并签字。

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **本科生实习&毕业设计（论文）工作周报** | | | | | | | | |
| 学生姓名 | 李天宝 | 联系电话 | 15704600640 | | Email | turingmac@hotmail.com | | |
| 实习基地  (或实验室) | 上海骇咕赛信息科技有限公司 | | 基地导师 | 丁盛豪 | | | 校内导师 | 王宏志 |
| 报告日期 | 2016-09-18 | | 实习周次 | 第9周 | | | | |
| 报告区间 | 2016-09-12 到 2016-09-18 | | | | | | | |
| 工作重点 | 选择研究方向 | | | | | | | |
| 工作内容及进度 | 1. 确定研究方向的数据集成 2. 了解公司现有数据库的情况 | | | | | | | |
| 遇到问题 | 1. 对于数据集成方面没有经验 | | | | | | | |
| 自我小结 | 本周将题目确定为数据集成方向，之前没有过相关工作，这个领域对于自己来说是比较新鲜的，应该会具有一定的挑战性 | | | | | | | |
| 下周计划 | 了解选题背景 | | | | | | | |
| 导师意见 | 签字： | | | | | | | |

**说明：**

1. 上述表格大小可以根据文字多少调整；
2. 每周要写工作周报并发送给校内导师，实习基地导师可根据需要发送；
3. 校内实习学生每2周要书面报呈周报给导师，当面指导并签字。

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **本科生实习&毕业设计（论文）工作周报** | | | | | | | | |
| 学生姓名 | 李天宝 | 联系电话 | 15704600640 | | Email | turingmac@hotmail.com | | |
| 实习基地  (或实验室) | 上海骇咕赛信息科技有限公司 | | 基地导师 | 丁盛豪 | | | 校内导师 | 王宏志 |
| 报告日期 | 2016-09-25 | | 实习周次 | 第10周 | | | | |
| 报告区间 | 2016-09-19 到 2016-09-25 | | | | | | | |
| 工作重点 | 了解选题背景 | | | | | | | |
| 工作内容及进度 | 1. 了解数据集成背景 2. 了解数据集成相关的现有工具 | | | | | | | |
| 遇到问题 |  | | | | | | | |
| 自我小结 | 本周对于数据集成领域最基础的内容进行一些了解，包括数据集成的应用范围、背景以及已存在的解决方案。尽管现有oracle等公司的数据集成工具，但是仍需开发出一些功能来集成到公司的环境中 | | | | | | | |
| 下周计划 | 验证选题的可行性 | | | | | | | |
| 导师意见 | 签字： | | | | | | | |

**说明：**

1. 上述表格大小可以根据文字多少调整；
2. 每周要写工作周报并发送给校内导师，实习基地导师可根据需要发送；
3. 校内实习学生每2周要书面报呈周报给导师，当面指导并签字。

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **本科生实习&毕业设计（论文）工作周报** | | | | | | | | |
| 学生姓名 | 李天宝 | 联系电话 | 15704600640 | | Email | turingmac@hotmail.com | | |
| 实习基地  (或实验室) | 上海骇咕赛信息科技有限公司 | | 基地导师 | 丁盛豪 | | | 校内导师 | 王宏志 |
| 报告日期 | 2016-10-02 | | 实习周次 | 第11周 | | | | |
| 报告区间 | 2016-09-26 到 2016-10-02 | | | | | | | |
| 工作重点 | 验证选题的可行性 | | | | | | | |
| 工作内容及进度 | 1. 验证选题可行性 2. 确定工作范围 | | | | | | | |
| 遇到问题 | 1. 数据集成范围过广，工作量太大，难以完成 | | | | | | | |
| 自我小结 | 本周基于对数据集成的了解和mentor进行了相关的讨论，一致认为数据集成的工作量过大，在整个实习周期内可能没有较好的效果，因此应在数据集成领域内确定具体的子选题 | | | | | | | |
| 下周计划 | 确定具体选题方向 | | | | | | | |
| 导师意见 | 签字： | | | | | | | |

**说明：**

1. 上述表格大小可以根据文字多少调整；
2. 每周要写工作周报并发送给校内导师，实习基地导师可根据需要发送；
3. 校内实习学生每2周要书面报呈周报给导师，当面指导并签字。

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **本科生实习&毕业设计（论文）工作周报** | | | | | | | | |
| 学生姓名 | 李天宝 | 联系电话 | 15704600640 | | Email | turingmac@hotmail.com | | |
| 实习基地  (或实验室) | 上海骇咕赛信息科技有限公司 | | 基地导师 | 丁盛豪 | | | 校内导师 | 王宏志 |
| 报告日期 | 2016-10-09 | | 实习周次 | 第12周 | | | | |
| 报告区间 | 2016-10-03 到 2016-10-09 | | | | | | | |
| 工作重点 | 确定具体选题方向 | | | | | | | |
| 工作内容及进度 | 1. 确定选题为模式集成 2. 了解数据库模式的相关特性 | | | | | | | |
| 遇到问题 | 1. 公司内的数据库模式差异性比较小 2. 需要获取更多不同的数据库模式 | | | | | | | |
| 自我小结 | 本周将工作确定在数据库表的模式集成上，这是数据集成的基础也是很有挑战性的一项任务，需要接下来阅读相关论文了解这个工作 | | | | | | | |
| 下周计划 | 阅读模式集成相关概述论文 | | | | | | | |
| 导师意见 | 签字： | | | | | | | |

**说明：**

1. 上述表格大小可以根据文字多少调整；
2. 每周要写工作周报并发送给校内导师，实习基地导师可根据需要发送；
3. 校内实习学生每2周要书面报呈周报给导师，当面指导并签字。

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **本科生实习&毕业设计（论文）工作周报** | | | | | | | | |
| 学生姓名 | 李天宝 | 联系电话 | 15704600640 | | Email | turingmac@hotmail.com | | |
| 实习基地  (或实验室) | 上海骇咕赛信息科技有限公司 | | 基地导师 | 丁盛豪 | | | 校内导师 | 王宏志 |
| 报告日期 | 2016-10-16 | | 实习周次 | 第13周 | | | | |
| 报告区间 | 2016-10-10 到 2016-10-16 | | | | | | | |
| 工作重点 | 阅读模式集成相关概述论文 | | | | | | | |
| 工作内容及进度 | 1. 阅读 Dong X L, Srivastava D. Big data integration[C]//Data Engineering (ICDE), 2013 IEEE 29th International Conference on. IEEE, 2013: 1245-1248. | | | | | | | |
| 遇到问题 | 1. 模式集成的过程需要确定将什么样的属性聚合在一起，这个衡量标准不好确定 | | | | | | | |
| 自我小结 | 本周主要阅读了 Dong X L, Srivastava D关于 big data integration的概述论文，对即将要做的工作有了一定的了解 | | | | | | | |
| 下周计划 | 阅读模式集成相关论文 | | | | | | | |
| 导师意见 | 签字： | | | | | | | |

**说明：**

1. 上述表格大小可以根据文字多少调整；
2. 每周要写工作周报并发送给校内导师，实习基地导师可根据需要发送；
3. 校内实习学生每2周要书面报呈周报给导师，当面指导并签字。

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **本科生实习&毕业设计（论文）工作周报** | | | | | | | | |
| 学生姓名 | 李天宝 | 联系电话 | 15704600640 | | Email | turingmac@hotmail.com | | |
| 实习基地  (或实验室) | 上海骇咕赛信息科技有限公司 | | 基地导师 | 丁盛豪 | | | 校内导师 | 王宏志 |
| 报告日期 | 2016-10-23 | | 实习周次 | 第14周 | | | | |
| 报告区间 | 2016-10-17 到 2016-10-23 | | | | | | | |
| 工作重点 | 阅读词间距离相关论文 | | | | | | | |
| 工作内容及进度 | 1. 将形式上相近作为属性类似的一个标准 2. 阅读 Levenshtein V I. Binary codes capable of correcting deletions, insertions and reversals[C]//Soviet physics doklady. 1966, 10: 707. 3. 阅读 L. Li, H. Wang, J. Li, and H. Gao. Ed-sjoin; an optimal algorithm for similarity oins with edit distance constraints [j]. Journal of Computer Research and Development, 46:319-325, 2009 | | | | | | | |
| 遇到问题 | 1. 判断两个属性相似的方式往往是对于两个词设计一个距离，相关的定义方式有很多，不好选择 | | | | | | | |
| 自我小结 | 本周主要确定衡量属性相似的一个维度，即形式上相近，这个情况往往发生在属性拼写错误，是一个很常见的情况，目前打算使用编辑距离来处理 | | | | | | | |
| 下周计划 | 阅读模式集成相关论文 | | | | | | | |
| 导师意见 | 签字： | | | | | | | |

**说明：**

1. 上述表格大小可以根据文字多少调整；
2. 每周要写工作周报并发送给校内导师，实习基地导师可根据需要发送；
3. 校内实习学生每2周要书面报呈周报给导师，当面指导并签字。

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **本科生实习&毕业设计（论文）工作周报** | | | | | | | | |
| 学生姓名 | 李天宝 | 联系电话 | 15704600640 | | Email | turingmac@hotmail.com | | |
| 实习基地  (或实验室) | 上海骇咕赛信息科技有限公司 | | 基地导师 | 丁盛豪 | | | 校内导师 | 王宏志 |
| 报告日期 | 2016-10-30 | | 实习周次 | 第15周 | | | | |
| 报告区间 | 2016-10-24 到 2016-10-30 | | | | | | | |
| 工作重点 | 阅读知识库相关论文 | | | | | | | |
| 工作内容及进度 | 1. 将语义上相近作为属性类似的一个标准 2. 阅读 Bollacker K, Evans C, Paritosh P, et al. Freebase: a collaboratively created graph database for structuring human knowledge[C]//Proceedings of the 2008 ACM SIGMOD international conference on Management of data. ACM, 2008: 1247-1250. 3. 阅读 ACM, 2014: 601-610. Dong, Xin, et al. "Knowledge vault: A web-scale approach to probabilistic knowledge fusion." 4. 阅读 Wu W, Li H, Wang H, et al. Probase: a probabilistic taxonomy for text understanding[C]//Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data. ACM, 2012: 481-492. | | | | | | | |
| 遇到问题 | 1. 现有的知识库不太容易判定什么样的两个属性在知识库中是应该归并为一条属性的 | | | | | | | |
| 自我小结 | 本周主要确定衡量属性相似的一个维度，即含义相近，多数情况是同义的两个属性。而这种判定应该符合人们的思维，因此我们采用web上手机的知识库来提供相关的信息 | | | | | | | |
| 下周计划 | 阅读模式集成相关论文 | | | | | | | |
| 导师意见 | 签字： | | | | | | | |

**说明：**

1. 上述表格大小可以根据文字多少调整；
2. 每周要写工作周报并发送给校内导师，实习基地导师可根据需要发送；
3. 校内实习学生每2周要书面报呈周报给导师，当面指导并签字。

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **本科生实习&毕业设计（论文）工作周报** | | | | | | | | |
| 学生姓名 | 李天宝 | 联系电话 | 15704600640 | | Email | turingmac@hotmail.com | | |
| 实习基地  (或实验室) | 上海骇咕赛信息科技有限公司 | | 基地导师 | 丁盛豪 | | | 校内导师 | 王宏志 |
| 报告日期 | 2016-11-06 | | 实习周次 | 第16周 | | | | |
| 报告区间 | 2016-10-31 到 2016-11-06 | | | | | | | |
| 工作重点 | 阅读知识库相关论文 | | | | | | | |
| 工作内容及进度 | 1. 探究其他判定属性类似的标准 2. 阅读 A. Arasu, S. Chaudhuri, and R. Kaushik. Learning string transformations from examples. Proceedings of the VLDB Endowment, 2(1):514{525, 2009. 1247-1250. 3. 阅读 R. Cilibrasi and P. Vitanyi. Automatic meaning discovery using google. In Dagstuhl Seminar Proceedings. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2006. 4. 阅读 U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth. From data mining to knowledge discovery in databases. AI magazine, 17(3):37, 1996. | | | | | | | |
| 遇到问题 | 1. 类似于Google距离或者 transformation学习的方式比较有效，但是所需要的时间、样本、人力的成本太高 | | | | | | | |
| 自我小结 | 本周主要了解其他可行的判定方式，由于开销太大，目前不作考虑 | | | | | | | |
| 下周计划 | 撰写开题报告 | | | | | | | |
| 导师意见 | 签字： | | | | | | | |

**说明：**

1. 上述表格大小可以根据文字多少调整；
2. 每周要写工作周报并发送给校内导师，实习基地导师可根据需要发送；
3. 校内实习学生每2周要书面报呈周报给导师，当面指导并签字。

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **本科生实习&毕业设计（论文）工作周报** | | | | | | | | |
| 学生姓名 | 李天宝 | 联系电话 | 15704600640 | | Email | turingmac@hotmail.com | | |
| 实习基地  (或实验室) | 上海骇咕赛信息科技有限公司 | | 基地导师 | 丁盛豪 | | | 校内导师 | 王宏志 |
| 报告日期 | 2016-11-13 | | 实习周次 | 第17周 | | | | |
| 报告区间 | 2016-11-07 到 2016-11-13 | | | | | | | |
| 工作重点 | 撰写开题报告 | | | | | | | |
| 工作内容及进度 | 1. 整合现有准备内容 2. 撰写开题报告 | | | | | | | |
| 遇到问题 | 1. 对部分细节了解不深入 | | | | | | | |
| 自我小结 | 本周主要将之前几周的研究内容进行综合，设计整个模式集成系统的框架，在撰写开题报告的过程中贵系统进行完善 | | | | | | | |
| 下周计划 | 准备开题答辩 | | | | | | | |
| 导师意见 | 签字： | | | | | | | |

**说明：**

1. 上述表格大小可以根据文字多少调整；
2. 每周要写工作周报并发送给校内导师，实习基地导师可根据需要发送；
3. 校内实习学生每2周要书面报呈周报给导师，当面指导并签字。

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **本科生实习&毕业设计（论文）工作周报** | | | | | | | | |
| 学生姓名 | 李天宝 | 联系电话 | 15704600640 | | Email | turingmac@hotmail.com | | |
| 实习基地  (或实验室) | 上海骇咕赛信息科技有限公司 | | 基地导师 | 丁盛豪 | | | 校内导师 | 王宏志 |
| 报告日期 | 2016-11-20 | | 实习周次 | 第18周 | | | | |
| 报告区间 | 2016-11-14 到 2016-11-20 | | | | | | | |
| 工作重点 | 准备开题答辩 | | | | | | | |
| 工作内容及进度 | 1. 完善开题报告 2. 做开题答辩PPT | | | | | | | |
| 遇到问题 |  | | | | | | | |
| 自我小结 | 本周主要准备开题答辩，未进行更多的工作 | | | | | | | |
| 下周计划 | 寻找并预处理知识库 | | | | | | | |
| 导师意见 | 签字： | | | | | | | |

**说明：**

1. 上述表格大小可以根据文字多少调整；
2. 每周要写工作周报并发送给校内导师，实习基地导师可根据需要发送；
3. 校内实习学生每2周要书面报呈周报给导师，当面指导并签字。