

综合论文训练记录表

学生姓名	丘骏鹏	学号	2009011282	班级	计92
论文题目	大数据环境下信息抽取模板自动聚类发现				
主要内容以及进度安排	<p>主要内容:</p> <p>目前已经获得海量的新闻、博客等原始网页数据,不同的网页采用不同的模板,我们希望能从数据的冗余性中发现网页模板,并用模板来抽取数据。</p> <p>系统包括几个部分:网页过滤、网页聚类和模板生成</p> <p>进度安排</p> <p>1-4周:任务分析,文献阅读,研究算法</p> <p>5-8周:系统的初步实现</p> <p>9-12周:系统改进和修正</p> <p>13-16周:毕业论文撰写</p> <p>指导教师签字: <u>朱立 黄民强</u></p> <p>考核组组长签字: <u>李建民</u></p> <p>2013年 3 月 21日</p>				
中期考核意见	<p>丘骏鹏同学工作努力,进展顺利,实现了开题时预定的目标。</p> <p>考核组组长签字: <u>李建民</u></p> <p>2013年 4 月 23日</p>				