**武汉理工大学本科生毕业设计(论文)任务书**

题目：基于CRF算法的中文社交媒体的事件发掘研究与实现

一、主要研究或设计内容 (300字左右)：

网络社会和现实社会处于同一时空维度下，与现实社会息息相关。随着网络社会的发展，它能越来越真实地反映出现实社会的当前关注点、指定方向关注点等社会信息，甚至对现实社会正在施以越来越大的影响力。社交媒体（Social Media）承载了大量的信息，从中可以发掘出很多有指导、启发意义的内容。

社交媒体的事件发掘在几年前便开始受到关注，大量以Twitter为代表的英文社交媒体事件发掘系统得到研究。然而由于中文的特殊性，例如表意方式多种、理解方式多种、缩略语及新生词较多等，中文社交媒体事件发掘中很多语言处理问题，例如分词、词性标注等相对英文来说具有更大的挑战性，已有的中文自然语言处理技术并不完全适合直接用于中文社交媒体信息的处理，精确度（Accuracy）和召回（Recall）都比较低。

基于以上，本课题希望能够将中文自然语言处理技术与社交媒体自身特性相结合，进行中文社交媒体的事件发掘研究，设计并实现基于CRF的中文社交媒体事件发掘工具。

二、完成的主要任务及要求：

主要任务：

1. 微博信息获取和中文语言处理工具研究；
2. CRF方法的原理的研究；
3. 实现微博事件发掘系统
4. 达到实现基于CRF的中文社交媒体事件发掘工具的目的

任务要求：

1. 查找相关文献；

2. 掌握CRF方法的原理，并掌握其在中文事件发掘上的应用方法；

3. 实现一个基于CRF中文社交媒体事件发掘工具；

4. 查阅参考文献15篇以上（其中近5年外文文献不少于3篇），参考文献撰写格式必须规范（按照《武汉理工大学本科生毕业设计（论文）撰写规范》）；

5. 完成不少于12000字的毕业论文撰写并完成答辩的相关工作；

6. 完成不低于5000汉字（20000英文印刷符）的教师指定的相关文献的英译汉翻译；

7. 正文应包含不少于12幅图（包括：原理图、流程图、结构框图、程序框图等）；

三、完成任务的时间节点：

第1周—第4周 搜集资料，撰写开题报告；

第5周—第6周 论文开题；

第7周—第12周 撰写论文初稿；

第13周—第16周 修改论文；

第17周 论文答辩。