**附件三：《2016年第六届“华为杯”中国大学生智能设计竞赛参赛作品报告》**

**作品名称：** 英文电子邮件写作助手

**参赛学校：** 中国石油大学(华东)

**学院/系：** 计算机与通信工程学院

**指导教师：** 曹绍华

**组 长：** 张斌

**组 员：** 王思宇 李祥璞

**通信地址：** 山东省青岛市黄岛区长江西路66号

**电 话：** 17854227061

**电子邮箱：** fealie28@163.com

**提交日期：** 2016年5月30日

填写说明

1. 所有参赛项目必须为一个基本完整的设计。参赛作品简介旨在能够清晰准确地阐述（或图示）该参赛队的参赛项目（或方案）。

2. 参赛作品简介采用A4纸撰写。除标题外，所有内容必需为宋体、小四号字、1.5倍行距。

3. 参赛作品简介不超过6页A4纸。

4. 参赛作品简介模板里已经列的内容仅供参考，作者也可以多加内容。

1. **摘要（请简要说明创作本作品之动机、功能、特性、创新处、实用性）**

本作品为设计一个系统，在用户输入英文的同时，能自动建议若干个提示，提示的结果可以有多个，并且提示的结果可以根据用户的使用情况进行调整，使得排在前面的提示结果被用户选用的可能性大。这个插件旨在提高人们书写英文电子邮件的效率，提示的选项会接近用户的写作习惯，使用户在写邮件的时候能够更多的关注邮件内容；对于不擅长英语写作的用户，本系统通过单词信息的提示，能够帮助这些用户更好的完成英文邮件，同时避免出现用词不当或拼写错误的情况。

1. **作品介绍（建议包括：特色描述、背景分析、相关工作及应用市场分析等）**

本作品主要研究内容为对用户输入进行检测并给出相应英文单词的提示，同时记录用户的选择，建立算法模型，根据从用户输入得到的这些记录信息对用户的写作习惯进行分析，来调整提示信息，使得更有可能被用户使用或对用户有用的提示排在前面。在算法上，我们将通过对已有资料的分析，结合用户的输入和使用习惯得出用户可能需要的提示内容，不断更新数据库中相关的提示信息及提示信息出现的次序，使得提示信息更加符合用户需求。

1. **实现方案（建议包括系统方案、实现原理、硬件框图、软件流程、功能、指标、特色等）**

1.系统方案

本作品旨在能够在用户输入英文邮件时给用户下一单词的提示，用户可以对给出的提示单词进行选择来代替从键盘输入单词，节省用户书写邮件的时间。系统分为向用户显示的单词提示窗口和保存查找单词的词库，单词提示窗口能够显示可供选择的单词，并在用户选择后写入到文本中；词库为记录英文文本中单词后出现某个单词的频数的一个行列阵，词库分为资料词库和用户词库，资料词库为我们通过分析已有英文文本数据产生的词库，用户词库为记录用户在使用时产生的数据，当检索到用户输入完一个单词后，会分别从资料词库和用户词库中查找此单词后单词出现频数前五位的单词（也可能不足五个）， 再从这十个词中根据原词库中频率乘以权重比（用户词库权重比为1，资料词库中单词出现频数小于9时权重比为0.6，大于等于时为1）的积（权重），选择权重值前五的单词显示在用户的单词提示窗口中，并将用户对单词提示窗口中单词的选择和用户的输入记录在用户词库中。

2.系统流程：



3.系统指标

1. 提示窗口中的单词正确率大于30%
2. 系统运行时流畅，无明显的滞后
3. 系统能够根据用户的输入及时更新提示单词排列顺序及内容

4.系统功能

在用户输入英文文本时，预测用户下一个可能要输入的单词，并将这些单词以可能性由高到低的顺序排列显示在单词提示窗口中，供用户选择。

1. **性能测试（建议包括测试方案、测试环境搭建、测试设备、测试数据、结果分析等）**

1.测试环境：windows操作系统

2.测试设备：个人电脑

3.测试方案：选取网络上的英文电子邮件文本，分别涉及旅行、约会、表达歉意、求职、请求帮助等不同主题，向系统中逐单词输入邮件。若提示内容中有此单词在原邮件中的下一个单词，则记一次正确，每次输入单词记录总次数加一。正确次数与记录总次数的比值记为正确率，用多次输入邮件所得正确率的平均值来估算系统的正确率，作为评价系统的指标。

4.测试数据及结果（只选取了部分测试结果）：

① Fill me in. How can i eavesdrop?

I belong to a natgas discussion group on the internet. This is from one of the guys. Basically they are a bunch of gastraders from various firms ( a lot of producers, some industrials, some small shops and few i--banker types) I found the test to be mildly amusing. And since I had no idea what 'club no minors' was-- i was hoping for some insight from you guys. Fortunately, the lovely Ms. Shipos was able to fill me in.

提示准确率：33.96%

② Anyway, while occasionally garbage, the discussions do provide insight on what others are thinking of production, storage and other such matters. And when I was a marketer, I found a few leads. Finally, since a lot of the information revolves around gossip about a particular 'super trader' at the big ENE , I find it amusing that one of the most reserved and modest individuals I know is so talked about on the internet.

提示准确率：26.92%

③ My fellow managers and I are delighted that your work in the marketing field has been recognized this way and we join in sending you our very best wishes for the future.

Through the five years of working together with you, many of us well aware of how much you've contributed to the association between our two corporations. We are all looking forward to your trip to China next month when we will celebrate your advancement in a more formal way.

Again, congratulations to you, Mr. Haskel - good luck and good wishes on your new position as Director of Marketing Department.

提示准确率：27.43%

5.结果分析 : 系统测试得到的平均正确率为29.63%，可认为接近30%即系统设计的最低指标，系统流畅性和及时更新提示单词排列顺序及内容符合系统设计指标。通过查看测试过程，我们发现主要原因在于资料词库信息不够完善，一些常用搭配没有包含在内，大部分的地名或专有名词词组不能给予正确提示，此外我们发现部分提示不正确的原因在于没有正确判断语句所用时态或人称，导致提示的单词形式与原文不同。

6.实现功能：输入英文邮件文本时系统能给出具有一定准确性的提示内容。

1. **创新性（（本部分内容主要说明作品的创新性）**

本作品的创新之处在于词库的内容会随着用户的输入或对单词提示内容的选择及时进行更新，并对提示信息中单词的次序进行动态的排序，使系统给出的单词提示信息能更加符合用户的书写习惯。我们认为随着用户使用此系统的时间增长，提示单词的准确度会逐渐增加。因为资料词库中记录的单词信息来源于不同文本，为确保准确性，资料词库中的出现频数小于9的单词在计算权值时乘以0.6的折扣，而对于资料词库中出现频数大于等于9的单词，我们认为它的出现概率较为准确，在计算权值时不会有折扣。

1. **总结**

在用户输入英文邮件文本时本系统能给出具有一定准确性的提示内容。由于我们使用的训练资料语库的文本数量和质量有所不足，系统的正确率还有待提高。