北 京 邮 电 大 学
本科毕业设计（论文）任务书
Project Specification Form

学院 School	International School	专业 Programme	Telecommunications	班级Class	2013215110
学生姓名 Name	ZHENG Weihong	学号 BUPT student no	2013213217	学号 QM student no	130801991
设计（论文）编号 Project No.	IC_3217				
设计（论文）题目 Project Title	Fine-grained Sentiment Analysis of Product Comments Based on Product Features				
论文题目（中文）	基于产品特征的产品评论细粒度情绪分析				
题目分类 Scope	Implementation	Computer Software	Software		

主要任务及目标Main tasks and target:	By
Task 1: Write a program that can crawl the online reviews on e-commerce websites, and complete the preprocessing tasks.	15/12/2016
Task 2: Write a program using python machine learning algorithm to extract feature words and opinion words.	20/02/2017
Task 3: Build up a relationship between features and their corresponding evaluating words.	01/04/2017
Task 4: Analyze the outcome of the program and come up with the conclusion you summarize.	01/05/2017

Measurable outcomes
1) A program that can collect online reviews from an e-commerce platform (e.g.Tmall).
2) Feature words and opinion words extracted from comments of an exact product.
3) A report of the users attitudes toward the certain features of the product you choose.

主要内容Project description：

The users often express their opinions and attitude to the components, function and capacity of the product via the online reviews.The review text greatly affects other people's purchase decision. Thus, the research on users' sentiment orientation on the products in the review text is very significant. As the development of e-commerce industry, we find that users tend to pay more attention to the product features than the overall quality of a product. In order to let e-commerce vendors know better how other people think of their product, this project should analysis the users' sentiment orientation based on product features from the review text. To deal with Chinese documentation, the necessary preprocessing tasks are word segmentation, eliminating stopword and stemming. In the next step, you can use some feature extraction algorithm to recognize the character of a product(such as price,quality,after sale service). Besides, you need to reveal the users general attitude towards every character by extracting opinion words and combining the features and opinion words into bigrams. In this stage, you should implement machine learning models and write a computer program to process the texts. If you have more interest, you can reseach the results of different models and make a compare.
At last, this project need a explanation or report to summarize the result of the program and give some improvement advice to the users or vendors.

Project outline

Costumers place much emphasis on the reviews of a product when they plan to buy it, therefore there exists demand to provide costumers with a convenient way to summarise other buyers' attitudes to the features of a particular product such as its price, quality, after sale service and so on. This project is to research users' sentiment orientation towards a product through their comments and give a report.
For the data I need, a program which can crawl online reviews will be developed and it will be used to collect user comments of a specific product I choose representatively from an e-commerce platform like Tmall. Being collected, these data will experience a preprocessing to reach an ideal state for further operations that have some requirements to data, preprocessing includes word segmentation, eliminating stopword and stemming.
The next step is to learn about some feature extraction algorithms and use them to recognize the characters of the product and extract opinion words, if this step cannot process as well as expected then the preprocessing for data will be considered to be fixed. Furthermore, some machine learning models will be implemented to build up relationships between features and their corresponding comments as a form of bigrams. Until this point, by summarising these bigrams, users' general attitude towards each character can be revealed.
Finally, a report will be given to analyse the outcome and give some improvement advices to users and vendors.
In order to finish the task, the Python v2.7 as well as its some packets including scrapy, numpy, scipy, scikit-learn will be used to help to program.

What I expect to have working at the mid-term oral

Finish the crawler program and collect data from a representative product.
Preprocess the data to satisfy requirements of further operations.
Finish preliminary feature and opinion extraction.

	Nov		Dec		Jan		Feb		Mar		Apr		May	
Task 1: Write a program that can crawl the online reviews on e-commerce websites, and complete the preprocessing tasks														
Learn about crawler programming and develop a crawler to collect data	■	■												
Learn about preprocessing technologies and preprocess data		■	■											
Task 2: Write a program using python machine learning algorithm to extract feature words and opinion words														
Learn about some feature extraction algorithms				■	■									
Preliminary feature and opinion extraction, preprocess data again if necessary						■	■	■						
Task 3: Build up a relationship between features and their corresponding evaluating words														
Evaluate former extraction and complete it								■	■					
Build up relationships between features and opinion words									■	■				
Task 4: Analyze the outcome of the program and come up with the conclusion you summarize														
Analyse the outcome according to features and their comments and write a draft report											■	■		
Complete the report													■	