



## Final project

Add the following information at the header of your assignment:

```
/*****  
Name:  
Assignment id: AIAF  
Due Date: May. 27, 2012  
*****/
```

### News agent

Design an intelligent news agent, which read news from the Web and do the following tasks:

1) At training stage, input a set of news, classified into five categories:

Sports, Entertainment, Health, Business, Sci/Tech,

Suggested number of training news: 20 news articles for each category

Software interface:

Input: a file folder name contain all the training articles, articles names are:

1.txt, 2.txt, ..., 100.txt

Have a separate file class.txt to indicate the categories of each article,

Sports: S

Entertainment: E

Health: H

Business: B

Sci/Tech: T

Each row in class.txt represent the category of one news article:

e.g.,

1.txt, S

2.txt, E

3.txt, T

...

100.txt, T

Output: the model that you build for classification

2) At testing stage, input a new article, and ask the agent to tell what class it is

Testing 100 different news articles, and report the accuracy

Input: the name of a new article, e.g., test.txt

Output: Class: e.g., T

3) At testing stage, input a set of news articles, and ask the agent to cluster them into different groups

# AI

## Course Project



Dr. Jinlin Chen

NSB -A118  
Dept. of Computer science  
Queens College, CUNY  
Flushing, 11367, NY

Tel: 718-997-3497 (o)  
Fax: 718-997-3513  
E-mail: [jchen@cs.qc.edu](mailto:jchen@cs.qc.edu)

Use google news as a reference, input articles from at least five topics, each topic with at least five articles, and then check how many groups are appropriate.

Input: a file folder name contain all the testing articles, articles names are:

t1.txt, t2.txt, ...,

Output: groups.txt, e.g.,

Group 1: t1.txt, t30.txt, t35.txt

Group 2: t2.txt, t4.txt, ...

### Requirement:

- 1) Provide document describing your basic idea for classification and clustering;
- 2) Provide document describing your code design and implementation
- 3) Provide source code with clear explanation
- 4) Provide sample news collection and sample testing results
- 5) Provide report on your testing results and analysis on your results

### Note:

- 1) You can manually edit each news article so that only text information is included;
- 2) You can collaborate on news article collection to save time