

**LING115 Homework #9**  
**Due: October 29, 2010**

Instructions:

Create a hw9 directory under your home directory and make sure you save all your work in that directory.

1. [2 points] Explain what the following regular expressions mean:

- 1) `sjsu*`
- 2) `^sjsu`
- 3) `sjsu\`
- 4) `sjsu|spartan`
- 5) `sjsu.+spartan`
- 6) `r[aeiou]m`
- 7) `(r[aeiou]){2}m`
- 8) `r[^aeiou]m`

Save your write-up as `hw9.q1` and put it under `hw9` in your home-directory.

2. [4 points] Write a Python program that identifies the following types of strings in the standard input:

- 1) e-mail address (e.g. `this.is@email.com`)
- 2) zip code (e.g. `95192`)
- 3) phone number (e.g. `408-123-4567`)
- 4) dollar amount (e.g. `$1234.56`)

Test your program with `normalize_me.txt` under `/home/ling115/hw9_out/`. The output of your program should look something like the following:

```
email: i.own.the.house@paymenrent.come, ioweunothing@renter.com
zip: 95777
phone: 408-123-4567, (215)-987-6543
dollar: $1234.56, $ 2469.02, $2459.12, $1
```

Save your program as `tokenizer.py` and put it under your `hw9` directory.