

# Processing text with regular expressions

Methods in Computational Linguistics I

# Last Time

- Introduction to Regular Expressions

# Today

- Processing Text with Regular Expressions

# Uses of Regular Expressions

- Cleaning Documents -- Isolating areas of interest
- Extracting information from texts or strings
- Converting from one format to another

# Email example

from Student Name <[student@domain.com](mailto:student@domain.com)>

date Wed, Oct 27, 2010 at 7:34 AM

subject Re: CS700 hw6 Grades

mailed-by gmail.com

Here you go.

On Wed, Oct 27, 2010 at 12:25 AM, Professor Name  
<[prof@domain.com](mailto:prof@domain.com)> wrote:

Please send your homework now.

# Demos

- Isolate the email text.
- Identify the Sender's email address
- Identify the date sent
  - Date parsing in python
- Anonymizing email addresses

# Tasks for today

- Combining file input and output with regular expressions
- Regular expression matching
- Regular expression replacement

# Next Time

- Part of speech tagging with NLTK.