LING/C SC/PSYC 438/538

Lecture 3

Sandiway Fong
Today’s Topics

• Homework 4 out
  • due next Tuesday by midnight
  • Homework 3 should have been submitted yesterday
• Quick Homework 3 review
• Continue with Perl intro
Homework 3 Review

Fortune Magazine’s Elon Musk AI article...

Sample student one-line summaries:

- Elon Musk is afraid of AI because it may eventually be able to outsmart us.
- Elon Musk, world renowned businessman, warns that artificial intelligence (AI) used to replace human roles could lead to our undoing as they act without concern for human safety in the pursuit of their tasks.

Notes:
- URL didn’t work for me.
- I did cut and paste...
Homework 3 Review


Headline:
For the first time ever, two Asian schools beat Yale in one popular university ranking

In the company’s just-released 2015/16 rankings, the National University of Singapore went from 22nd last year to 12th this year, and Nanyang Technological University went from 39th to 13th. The schools—both in Singapore—became the first in Asia to break into the top 15. Both placed ahead of Yale, which fell from 10th to 15th.

5% summarization = table + caveat

<table>
<thead>
<tr>
<th>2015/16 rank</th>
<th>University 2014/15 rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Massachusetts Institute of Technology (MIT) 1</td>
</tr>
<tr>
<td>2</td>
<td>Harvard University 4</td>
</tr>
<tr>
<td>3</td>
<td>University of Cambridge 2</td>
</tr>
<tr>
<td>4</td>
<td>Stanford University 7</td>
</tr>
<tr>
<td>5</td>
<td>California Institute of Technology 8</td>
</tr>
<tr>
<td>6</td>
<td>University of Oxford 5</td>
</tr>
<tr>
<td>7</td>
<td>University College London 5</td>
</tr>
<tr>
<td>8</td>
<td>Imperial College London 2</td>
</tr>
<tr>
<td>9</td>
<td>Swiss Federal Institute of Technology 12</td>
</tr>
<tr>
<td>10</td>
<td>University of Chicago 11</td>
</tr>
<tr>
<td>11</td>
<td>Princeton University 9</td>
</tr>
<tr>
<td>12</td>
<td>National University of Singapore 22</td>
</tr>
<tr>
<td>13</td>
<td>Nanyang Technological University 39</td>
</tr>
<tr>
<td>14</td>
<td>Ecole Polytechnique Fédérale de Lausanne 17</td>
</tr>
<tr>
<td>15</td>
<td>Yale University 10</td>
</tr>
</tbody>
</table>
Clickbait articles are syndicated?

A juice company dumped orange peels in a national park. Here
www.upworthy.com/a-juice-company-dumped-orange-peels-in-a-national-park-heres... ▼
5 days ago - Here's what it looks like now. ... 16 years later, Janzen dispatched graduate student
Timothy Treuer to look for the site where the food waste ...

Juice company dumped orange peels in deforested area. Here's what ...
shareably.net/.../juice-company-dumped-orange-peels/ ▼
4 days ago - Here's what it looks like 16 years later. Share. Juice company dumped orange peels in
deforested area. Here's what it looks like 16 years later ...

Juice company dumped orange peels in deforested area. Here's what ...
www.trendolizer.com/.../juice-company-dumped-orange-peels-in-deforested-area-her... ▼
2 days ago - The juice company dumped 12000 tons of orange peels in the area, and were stunned
when they ... Here's what it looks like 16 years later.

Juice company dumped orange peels in deforested area. Here's what ...
https://www.reddit.com/.../juice_company_dumped_orange_peels_in_deforested/ ▼
2 days ago - When discussing specific recycling in an area, please mention the country - IE [ UK]. ...
Here's what it looks like 16 years later (shareably.co).

A juice company dumped orange peels in a national park. Here's what ...
creativespk.com/.../a-juice-company-dumped-orange-peels-in-a-national-park-heres... ▼
5 days ago - A juice company dumped orange peels in a national park. Here's what it looks like now.
... orange peels and pulp, free of charge, in a heavily grazed, largely deforested area nearby. ... 16
years later, Janzen dispatched graduate student Timothy Treuer to look for the site where the food
waste was dumped.
Homework 3 Review

How does OTS work?
• The Idea behind OTS is that the important ideas in an article are described with many of the same words while redundant information uses less technical terms and is not related to the main subject of the article. Important lines are lines that are related to the subject of the article. The subject of the article is the list of ideas that are most discussed in the article.
• https://github.com/neopunisher/Open-Text-Summarizer
Homework 3 Review

**Step 1:** Frequency information (unigram):
- 11 are
- 17 is
- 16 a
- 14 Harry
- 14 on
- 13 Sally
- 11 Love
- 11 such
- 4 an
- 2 taxi
- 1 university
- 1 chicago
- 1 meets...

**Step 2:** Remove stopwords:
- 14 Harry
- 13 Sally
- 11 Love
- 2 taxi
- 1 university
- 1 chicago
- 1 meets...

**Then:**

1. Grade each sentence according to the wordlist.
2. Match will include **stemming**:
   - e.g. *run, running, runnable*
   - fold plural/singular
3. Pick the N% top-scoring sentences
Homework 4

**Question 1:** JM Section 1.2 talks about Ambiguity.

- Go to the Stanford University Parser
- Try the sentence from section 1.2: *I made her duck.*
- Analyze the resulting parse.
- Explain which of the interpretations (1.5) to (1.9) is the parse obtained by the system compatible with?
Homework 4 Hints

• **How to read the parse**
  • [http://languagelog.ldc.upenn.edu/myl/PennTreebank1995.pdf](http://languagelog.ldc.upenn.edu/myl/PennTreebank1995.pdf)

• **Phrasal Labels:**
  • VP = Verb Phrase, S = Sentence, NP = Noun Phrase, PP = Prepositional Phrase. ADJP = Adjectival Phrase.

• **Part of Speech (POS) Labels:**
  • DT = Determiner, JJ = Adjective, JJR = adjective (Comparative Form), NN = common noun, VB = verb (base form), IN = Preposition, PRP$ = Possessive Pronoun.
  • More here: [https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html](https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html)
Homework 4

**Question 2**: JM Section 1.6.3 talks about natural language understanding and the Blocks World simulation:

- *Move the red block on top of the smaller green one*

- Give two different possible structural readings for the sentence
  - Give parses with (...); i.e. in the style of the Stanford Parser
    - put in as much detail as you think you need to make the difference clear
  - **Hint**: think about what the (multiword) preposition *on top of* can modify
  - Explain your parses

- Use the Stanford Parser on the sentence:
  - which one do you think the parse most closely represents? Explain.
Last Time: Perl Print

```perl
1 print "Hello World";
2 print "Goodbye World\n"
```

```python
1 print("Hello World", end=' ')
2 print("Goodbye World")
```

```bash
<table>
<thead>
<tr>
<th>dhcp-10-142-178-220:Desktop sandiway$ perl test.perl</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hello WorldGoodbye World</td>
</tr>
</tbody>
</table>
```

```bash
<table>
<thead>
<tr>
<th>dhcp-10-142-178-220:Desktop sandiway$ python3 test.py</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hello WorldGoodbye World</td>
</tr>
</tbody>
</table>
```

Same behavior, different assumptions about end of line (\n) from these two programming languages

Note also the differences in syntax: (..) in Python3, no parentheses in Perl
; need to separate the statements in Perl, none in Python
Perl intro

- [Link](http://perldoc.perl.org/perlintro.html)

  1. `#!/usr/bin/perl`
  2. `use strict;`
  3. `use warnings;`

  "Nanny mode": helps with debugging

- **Scalars**: variables begin with ($)
  - no such type requirement in Python

- **Numbers (integer, floating point)**
  - Python includes complex numbers (cmath library)

- **Strings**

- **References** (pointers to (non-)scalars)
Perl intro

Notes from the tutorial:

• semicolon (;) is not always necessary
  • Command separator semantics vs. end of command (termination) token
  • Best practice? Terminate every command with a semicolon

• Variable types:
  • Every variable type has its own namespace.
  • This means that $foo and @foo are two different variables.
  • It also means that $foo[1] is a part of @foo, not a part of $foo. This may seem a bit weird, but that's okay, because it is weird.
Perl Arrays

*like a simple ordered list...*

- **Literal:**
  - `@ARRAY = ( ... , ... , ...)` (round brackets; comma separator)

- **Access:**
  - `$ARRAY[ INDEX]` (zero-indexed; negative indices ok; slices ok)

- **Index of last element:**
  - `@#array` (a scalar)

- **Coercion:**
  - `@ARRAY = number of elements in scalar context`

- **Built-in functions:**
  - `sort @ARRAY; reverse @ARRAY; push @ARRAY, $ELEMENT; pop @ARRAY; shift @ARRAY; unshift @ARRAY, $ELEMENT; splice @ARRAY, $OFFSET, $LENGTH, $ELEMENT`
  - `$ELEMENT above can be @ARRAY`

- **Built-in arrays:**
  - `@ARGV` (command line arguments)
  - `@_` (sub(routine) arguments)
Perl Variable Prefixation

**Reading Perl code (perldata) aloud:**

- *Larry Wall was a linguist* ...
- The '$_' symbol works semantically like the English word "the" in that it indicates a single value is expected.

- Entire arrays (and slices of arrays and hashes) are denoted by '@', which works much like the word "these" or "those" does in English, in that it indicates multiple values are expected.

Also:

- Entire hashes are denoted by '%'... (no translation)
- In addition, subroutines are named with an initial '&', though this is optional when unambiguous, just as the word "do" is often redundant in English.
Exercises

• If you’re new to programming, practice using Perl
  • read the intro and run all the examples!

• Preparatory reading:
  • Chapter 2 of the textbook
  • we’ll be using Perl’s regular expression (regex) engine for this