

LING/C SC/PSYC 438/538

Lecture 13

Prof. Sandiway Fong

Today's Topic

- Homework 8 Review
- Last Time:
 - two ways of inserting Perl code into regex
 - `s/regex/code/e`
 - `(?code)`
- More on powerful features in Perl regex:
 - lookahead
 - lookbehind
- Predicate-Argument Structure
 - Framenet
 - Stanford CoreNLP

Homework 8 Review

- **Format** (1st line):

- id¹,icij_sar_id²,filer_org_name_id³,filer_org_name⁴,**begin_date**⁵,**end_date**⁶,or
iginator_bank_id⁷,**originator_bank**⁸,originator_bank_country⁹,originator_iso¹
⁰,beneficiary_bank_id¹¹,**beneficiary_bank**¹²,beneficiary_bank_country¹³,bene
ficiary_iso¹⁴,**number_transactions**¹⁵,**amount_transactions**¹⁶

- **Notes:**

- "May 31, 2012","May 31, 2012"
- The Bank of New York Mellon Corp.
- "Citigroup, Inc."

Homework 8 Review

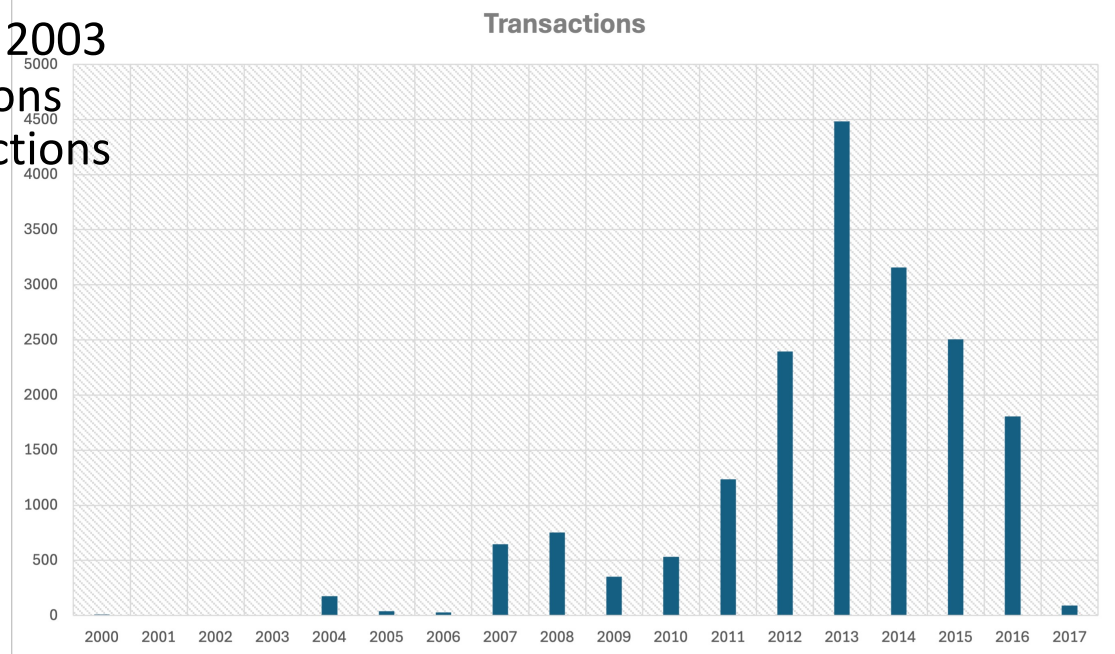
- Basic code:

- `perl -le 'open $f, $ARGV[0]; while(<$f){print "..."
if
/(".+?",|. *?,){4}"(. *?$ARGV[1])", ". *?", (" .+?",|. *?,)
{8}(. *?), (. *)/}' download_transactions_map.csv 2000`

Homework 8 Review

- **Question 1:**

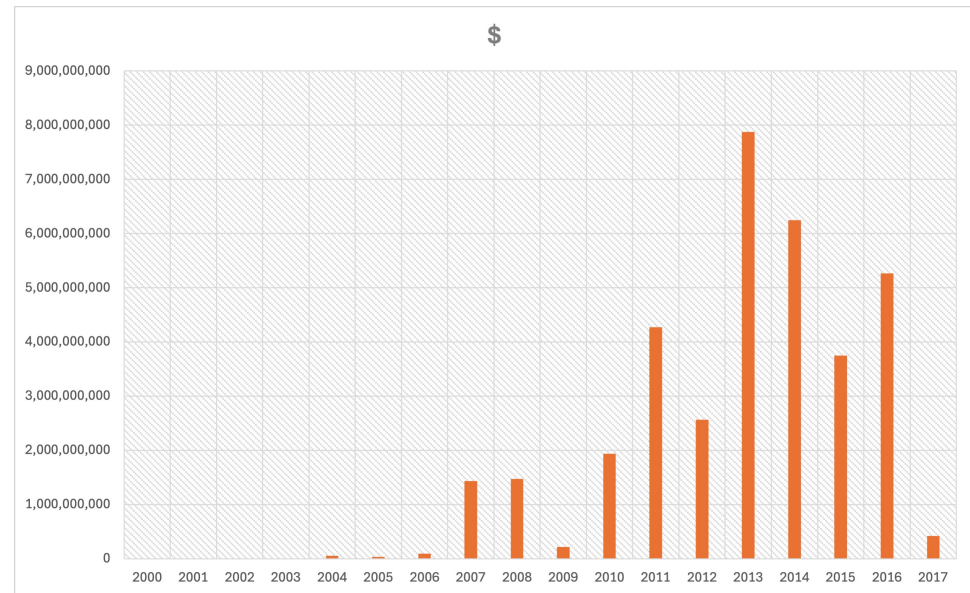
- (A) 2013 2014 2015 2012 2016 2011 2008 2007 2010 2009 2004 2017 2005
2006 2000 2002 2001 2003
- (B) 2003: no transactions
- (C) 2013: 4484 transactions



Homework 8 Review

- **Question 2:**

- (A) 2013 2014 2016 2011 2015 2012 2010 2008 2007 2017 2009 2006 2004
2005 2000 2002 2001 2003
- (B) 2013: \$7,871,446,356.98



Homework 8 Review

- **Question 3:** originator_bank⁸ by number_transactions¹⁵

- ```
perl -le 'open $f, $ARGV[0]; while(<$f>){if
(/(".*+?",|.*?,){7}(".*+?"|.*?),(".*+?",|.*?,){6}(.*)/) {if ($4)
{$c{$2}+=$4} else {$c{$2}++}}; for(keys %c){print "$c{$_} $_"}'
download_transactions_map.csv | sort -rn
```

478 AS Expobank

401 Western Union Business Solutions

377 Standard Chartered Bank

362 Primorye Bank

296 VP Bank AG

282 CIMB Bank Berhad

278 DBS Bank Ltd

278 Barclays Bank Plc Ho UK

253 Deutsche Bank AG

248 JSC Norvik Banka

# Homework 8 Review

- **Question 3:** originator\_bank<sup>8</sup> by **amount\_transactions**<sup>16</sup>

- ```
perl -le 'open $f, $ARGV[0]; while(<$f>){if
(/(".*+?",|.*?,){7}(".*+?"|.*?),(".*+?",|.*?,){7}(.+)$/) {$c{$2}+=$4}};
for(keys %c){print "$c{$_} $_"}' download_transactions_map.csv | sort
-rn
```

3157943202.12 Amsterdam Trade Bank NV

2666430134.11 AS Expobank

1702860761.18 Deutsche Bank AG

1375421435.08 ING Netherland NV

1366123530.96 Rigensis Bank AS

1207455547.63 Rosbank

1114866517.7 JPMorgan Chase Bank

926899916.66 Gazprombank

856543576.77 Caledonian Bank Ltd

760041526 Societe Generale Private Banking

Homework 8 Review

- **Question 4:** beneficiary_bank¹² by **number_transactions**¹⁵

```
• perl -le 'open $f, $ARGV[0]; while(<$f>){if  
  (/(".*?",|.*?){11}(".*?",|.*?){2}(.*)/,/){if ($4)  
    {$c{$2}+=$4} else {$c{$2}++}}; for(keys %c){print "$c{$_} $_"}'  
download_transactions_map.csv | sort -rn
```

- 742 Credit Suisse AG
- 480 JSC Norvik Banka
- 414 Barclays Bank Plc
- 385 Habib Bank A G Zurich
- 309 HSBC
- 294 Standard Chartered Bank
- 248 United Overseas Bank Ltd
- 237 Raiffeisen Bank International Ag
- 237 Deutsche Bank AG
- 215 United Overseas Bank Limited

Homework 8 Review

- **Question 4:** beneficiary_bank¹² by **amount_transactions**¹⁶
 - ```
perl -le 'open $f, $ARGV[0]; while(<$f>){if (/(".*+?"|.*?)(".*+?"|.*?)(".*+?"|.*?){3}(.+)$/) { $c{$2}+=$4}}; for(keys %c){print "$c{$_} $_"}' download_transactions_map.csv | sort -rn
```
- 3559645309.54 Rosbank
- 2089081782.75 Credit Suisse AG
- 1754253899.57 Rigensis Bank AS
- 1324026206.64 ING Netherland NV
- 1082748112.02 Deutsche Bank AG -- London Branch
- 881880017.38 JPMorgan
- 860707022.51 Societe Generale Bank And Trust Singapore SA
- 813175350.59 Hong Kong And Shanghai Banking Corp
- 805642190.61 Bank Soyuz
- 803999980 Caledonian Bank Ltd

# Regex Lookahead and Lookbehind

- We've already seen some **zero-width** regexs:
  - `^` (start of string)
  - `$` (end of string)
  - `\b` (word boundary)
    - matches the imaginary position between `\w\W` or `\W\w`, or just before beginning of string if `^\w`, just after the end of the string if `\w$`
- zero-width because position of match (so far), pos, doesn't change!
  1. `(?=regex)` (lookahead from current position)
  2. `(?<=regex)` (lookbehind from current position)
  3. `(?!regex)` (negative lookahead)
  4. `(?<!regex)` (negative lookbehind)

# Lookahead (and lookbehind)

## Lookaround Assertions

Lookaround assertions are zero-width patterns which match a specific pattern without including it in `$&`. Positive assertions match when their subpattern matches, negative assertions match when their subpattern fails. Lookbehind matches text up to the current match position, lookahead matches text following the current match position.

`(?=pattern)`

A zero-width positive lookahead assertion. For example, `/\w+(?=\t)/` matches a word followed by a tab, without including the tab in `$&`.

`(?!pattern)`

A zero-width negative lookahead assertion. For example `/foo(?!bar)/` matches any occurrence of "foo" that isn't followed by "bar". Note however that lookahead and lookbehind are NOT the same thing. You cannot use this for lookbehind.

If you are looking for a "bar" that isn't preceded by a "foo", `/(!foo)bar/` will not do what you want. That's because the `(!foo)` is just saying that the next thing cannot be "foo"--and it's not, it's a "bar", so "foobar" will match. Use lookbehind instead (see below).

`/(?<=\t)\w+/`

`(?<=pattern)` lookbehind for *pattern*

`(?<!pattern)` negative lookbehind for *pattern*

# Regex Lookahead and Lookbehind

- Example:

```
1 $s = "_bison _cat snake _dog cat _snake dog";
2 while ($s =~ /_(\w+)\b(?=.*\1\b)/g) {
3 print "<$1>\n"
4 }
```

```
$perl test.perl
<cat>
<dog>
$
```

looks for a word beginning with `_` such that there is a duplicate ahead (without the `_`)  
(`?= ..`) means **lookahead**

# Regex Lookahead and Lookbehind

## Some restrictions apply:

**lookbehind** (*in most versions of Perl*) **cannot** be of variable length

- From perlretut:
  - **Lookahead** can match arbitrary regexps,
  - but **lookbehind** prior to 5.30 (`?<=fixed-regexp`) only works for regexps of fixed width, *i.e.*, a fixed number of characters long. Thus `(?<=(ab|bc))` is fine, but `(?<=(ab)*)` prior to 5.30 is not.

# Debugging Perl regex

- (`{ Perl code }`) can be inserted anywhere in a regex
- can assist with debugging
- **Example:**

```
1 $s = "_bison _cat snake _dog cat _snake dog";
2 while ($s =~ /_(\w+)\b(?{print "$1\n"})(?=.*\1\b)/g) {
3 print "<$1>\n"
4 }
```

```
$perl test.perl
bison
cat
<cat>
dog
<dog>
snake
```

# Regex Lookahead and Lookback

- `(?<!pattern)`

- `(*nlb:pattern)`

- `(*negative_lookbehind:pattern)`

`/(?<!bar)foo/`

A zero-width negative lookbehind assertion. For example `/(?<!bar)foo/` matches any occurrence of "foo" that does not follow "bar".

Prior to Perl 5.30, it worked only for fixed-width lookbehind, but starting in that release, it can handle variable lengths from 1 to 255 characters as an experimental feature. The feature is enabled automatically if you use a variable length lookbehind assertion, but will raise a warning at pattern compilation time, unless turned off, in the `experimental::vlb` category. This is to warn you that the exact behavior is subject to change should feedback from actual use in the field indicate to do so; or even complete removal if the problems found are not practically surmountable. You can achieve close to pre-5.30 behavior by fatalizing warnings in this category.

# Background

- Background stuff you should familiar yourself with ...
  - Predicate-argument structure
  - Stanford CoreNLP

# Background

## Predicate-Argument Structure (typically for verbs)

- Example
  - *John saw/noticed the javelina*
  - *notice*(experiencer, theme) or *see*(experiencer, theme)
  - John noticed that Mary saw the javelina
  - *notice*(perceiver, proposition) 1<sup>st</sup> argument: subject, 2<sup>nd</sup> argument: direct object
  - the cat chased the mouse
  - *chase*(agent, theme) 1<sup>st</sup> argument: subject, 2<sup>nd</sup> argument: direct object
  - the mouse was chased by the cat (*passivization*)
  - \*John was jogged for an hour (\**passivization*)
  - John jogged for an hour (intransitive)
  - *jog*(agent)

# Background

- Different representations exist in the literature.
- Simple:
  - *John saw/noticed the javelina*
  - *notice*(experiencer, theme)
- Neo-Davidsonian:
  - event(*e*) & experiencer(*e*, John) & theme(*e*, javelina)

# Background

- Framenet
  - <https://framenet.icsi.berkeley.edu/luIndex>
  - Words in this frame have to do with a **Cognizer** adding some **Phenomenon** to their model of the world.

## Lexical Entry

**notice.v**

**Frame: Becoming\_aware**

**Definition:**

COD: become aware of.

### Frame Elements and Their Syntactic Realizations

The Frame Elements for this word sense are (with realizations):

| Frame Element     | Number Annotated | Realization(s)                                                         |
|-------------------|------------------|------------------------------------------------------------------------|
| <b>Cognizer</b>   | (25)             | CNI.-- (3)<br>NP.Ext (22)                                              |
| <b>Ground</b>     | (7)              | PP[from].Dep (1)<br>PP[in].Dep (2)<br>PP[at].Dep (3)<br>PP[on].Dep (1) |
| <b>Phenomenon</b> | (24)             | NP.Obj (16)<br>Sfin.Dep (7)<br>Sinterrog.Dep (1)                       |
| <b>State</b>      | (3)              | INI.-- (2)<br>VPing.Dep (1)                                            |
| <b>Time</b>       | (2)              | PP[on].Dep (1)<br>AVP.Dep (1)                                          |

core

core

# Background

## Framenet Examples:

- 420-that-sfin
  1. [CognizerI] soon **NOTICED** [Phenomenonthat the car was being driven very dangerously] .
  2. Then off they went but [CognizerI] had **NOTICED** [Phenomenonthat Mrs Taylor was really crying] .
  3. [CognizerYou] will **NOTICE** that there is , [Groundin the wording of that letter] , [Phenomenonsomething curious] .
- 430-sfin
  1. **NOTICE** [Phenomenonthe street names] [Groundin the centre of Bristol] .[CognizerCNI]
  2. [CognizerYou] may **NOTICE** [Phenomenonthat food tastes different when you are pregnant] .
  3. ` I do n't suppose [Cognizeranyone] will even **NOTICE** [Phenomenonyou 're not there] .
  4. [CognizerNobody] even **NOTICED** [PhenomenonI was in the room !]
- 480-swh
  1. On the way [Cognizerhe] **NOTICED** [Phenomenonhow quiet the school seemed] .
- 520-np-vping
  1. ` Did [Cognizeryou] **NOTICE** [Phenomenonany knives] [Statelying about] ? "
- 570-np-ppabout
  1. ` I see [Cognizeryou] have **NOTICED** [Phenomenona certain peculiarity about my appearance .] "
- 570-np-ppat
  1. When examining the wound , [CognizerI] **NOTICED** [Phenomenona dark area] [Groundat each end of the cut] .
  2. [CognizerUsers of the main car park at Park Royal] will have **NOTICED** [Phenomenona new fence] [Groundat the back of the site] .
  3. Then [CognizerI] **NOTICED** [PhenomenonAlec] [Groundat the other end of the bench] .

# Background

## Lexical Units:

- *chance (across).v, chance (on).v, come (across).v, come (upon).v, descry.v, detect.v, discern.v, discover.v, discovery.n, encounter.v, espy.v, fall (on).v, find (oneself).v, find out.v, find.v, happen (on).v, learn.v, locate.v, note.v, notice.v, observe.v, perceive.v, pick up.v, recognize.v, register.v, spot.v, spy out.v, tell.v*

## Not present **Perception\_experience** verbs:

- *detect.v, experience.n, experience.v, feel.v, hear.v, overhear.v, perceive.v, perception.n, see.v, sense.v, smell.v, taste.v, witness.v*

## Not present **Perception\_active** verbs:

- *admire.v, attend.v, eavesdrop.v, eye.v, feel.v, gape.v, gawk.v, gaze.n, gaze.v, glance.n, glance.v, goggle.v, listen.v, look.n, look.v, observation.n, observe.v, palpate.v, peek.n, peek.v, peep.v, peer.v, savour.v, smell.v, sniff.n, sniff.v, spy.v, squint.v, stare.n, stare.v, taste.n, taste.v, view.v, watch.v*

# Background

- Unified Verb Index

- <https://verbs.colorado.edu/verb-index/vn3.3/>

|                   |                                                                                                                |
|-------------------|----------------------------------------------------------------------------------------------------------------|
| notice            | <b>SEE-30.1-1-1, (PropBank), (FN BECOMING_AWARE), (GROUPING)</b>                                               |
| <b>FRAMES</b>     |                                                                                                                |
| <b>NP V S</b>     |                                                                                                                |
| EXAMPLE           | "I saw her bake the cake."                                                                                     |
| SYNTAX            | <u>EXPERIENCER</u> <b>V</b> <u>STIMULUS</u> <b>&lt;+OC_BARE_INF&gt;</b>                                        |
| SEMANTICS         | <b>PERCEIVE</b> (DURING(E), <b>EXPERIENCER</b> , <b>STIMULUS</b> ) <b>IN_REACTION_TO</b> (E, <b>STIMULUS</b> ) |
| <b>NP V S_ING</b> |                                                                                                                |
| EXAMPLE           | "I saw him laughing."                                                                                          |
| SYNTAX            | <u>EXPERIENCER</u> <b>V</b> <u>STIMULUS</u> <b>&lt;+OC_ING&gt;</b>                                             |
| SEMANTICS         | <b>PERCEIVE</b> (DURING(E), <b>EXPERIENCER</b> , <b>STIMULUS</b> ) <b>IN_REACTION_TO</b> (E, <b>STIMULUS</b> ) |
| <b>NP V S_ING</b> |                                                                                                                |
| EXAMPLE           | "I saw their laughing and joking."                                                                             |
| SYNTAX            | <u>EXPERIENCER</u> <b>V</b> <u>STIMULUS</u> <b>&lt;+POSS_ING&gt;</b>                                           |
| SEMANTICS         | <b>PERCEIVE</b> (DURING(E), <b>EXPERIENCER</b> , <b>STIMULUS</b> ) <b>IN_REACTION_TO</b> (E, <b>STIMULUS</b> ) |

# Background: Propbank

| <b>Roleset id:</b> <b>notice.01</b> , <i>become aware of</i> , <b>Source:</b> , vncls: , framnet:                                                                     |       |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|
| <b>notice.01:</b> NOTICE-V NOTES: Frames file for 'notice' based on sentences in wsj. Verb 30.1-1. Framed by Katie. (from notice.01-n) NOTICING-N, TAKE_NOTICE-L NOTE |       |
| <b>Aliases:</b>                                                                                                                                                       |       |
| Alias                                                                                                                                                                 | Frame |
| <b>notice</b> (v.)                                                                                                                                                    |       |
| <b>notice</b> (n.)                                                                                                                                                    |       |
| <b>noticing</b> (n.)                                                                                                                                                  |       |
| <b>take_notice</b> (l.)                                                                                                                                               |       |
| <b>Roles:</b>                                                                                                                                                         |       |
| <b>Arg0-PPT:</b> <i>noticer</i> (vnrole: 30.1-1-Experiencer)                                                                                                          |       |
| <b>Arg1-PAG:</b> <i>noticed</i> (vnrole: 30.1-1-Stimulus)                                                                                                             |       |

- Propbank:

- • ARGn-PAG ... proto-agent
- • ARGn-PPT ... proto-patient

# Background: Propbank

## notice-v; 2 Senses

- **Sense Number 1: observe, perceive or become aware of something**

- Examples:

Did you **notice** what he had in his hand?

I **noticed** that he avoided mentioning her name.

Mary waved at the man but he didn't seem to **notice**.

Starting in 1987, scientists **noticed** large drops in the amount of phytoplankton.

Her musical talent was first **noticed** by the critics at the age of 12.

- Mappings:

VerbNet: see-30.1-1-1

FrameNet: Becoming\_aware

PropBank: notice.01

WordNet 3.0 Sense Numbers: 1, 2, 4

# Background: Propbank

## notice-v; 2 Senses

- **Sense Number 2: bring to attention; give notice or announce**

- Examples:

The Solicitor General **noticed** the court of a change in Justice Department police.

The foundation **noticed** the Council of the new approach.

- Mappings:

VerbNet: NM

FrameNet: NM

PropBank: NM

# Background

## Predicate-Argument Structure (typically for verbs)

- Example
  - \*the librarian put the book
  - the librarian put the book on the table
  - *put*(agent, theme, location)
  - Mary gave John the textbook
  - \*Mary gave John
  - *give*(agent, goal, theme)
  - Mary gave the textbook to John

# Background: Framenet

## give:

|                                       |                                                                                                                          |
|---------------------------------------|--------------------------------------------------------------------------------------------------------------------------|
| <b>Core:</b>                          |                                                                                                                          |
| <b>Donor [Donor]</b>                  | The person that begins in possession of the <b>Theme</b> and causes it to be in the possession of the <b>Recipient</b> . |
| <b>Recipient [Rec]</b>                | The entity that ends up in possession of the <b>Theme</b> .                                                              |
| <b>Theme [Thm]</b>                    | The object that changes ownership.                                                                                       |
| <b>Semantic Type:</b> Physical_object |                                                                                                                          |

## put:

|                                                             |                                                                                                                                                                           |
|-------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Core:</b>                                                |                                                                                                                                                                           |
| <b>Agent [Agt]</b><br><b>Semantic Type:</b> Sentient        | The <b>Agent</b> is the person (or other force) that causes the <b>Theme</b> to move.<br><b>The waiter PLACED</b> the food on the table.                                  |
| <b>Cause [Cause]</b><br><b>Excludes:</b> Agent              | Grass , which is sown with clover , provides rich pasture for cattle in summer and the clover is <b>another plant which PUTS</b> nitrogen into the soil .                 |
| <b>Goal [Goal]</b><br><b>Semantic Type:</b> Goal            | The FE <b>Goal</b> is the location where the <b>Theme</b> ends up. This FE is profiled by words in this frame.<br>The waiter <b>PLACED</b> the food <b>on the table</b> . |
| <b>Theme [Thm]</b><br><b>Semantic Type:</b> Physical_object | The <b>Theme</b> is the object that changes location during the Placing.<br>The waiter <b>PLACED</b> <b>the food</b> on the table.                                        |