

LING/C SC/PSYC 438/538

Lecture 11

Prof. Sandiway Fong

Today's Topics

- Some comments on Homework 7
- *The lines of code that changed everything*

Getting deeper into Perl regex

- *word boundaries*
- *capture*
- *backreferences*
- *shortest vs. greedy matching*
- *nondeterminism (backtracking)*

Homework 7 Review

Unless otherwise noted, our bias ratings are based on **online, written content**, not TV or radio content.



FEATURED

ALL SOURCES

No results found for "courthouse news".

Courthouse News Service is rated as least biased with high factual reporting by [Media Bias Fact Check](#).

- **Political Bias:** Center / Least Biased
- **Factual Reporting:** High
- **Focus:** Legal and court reporting rather than partisan political commentary.

 Media Bias/Fact Check

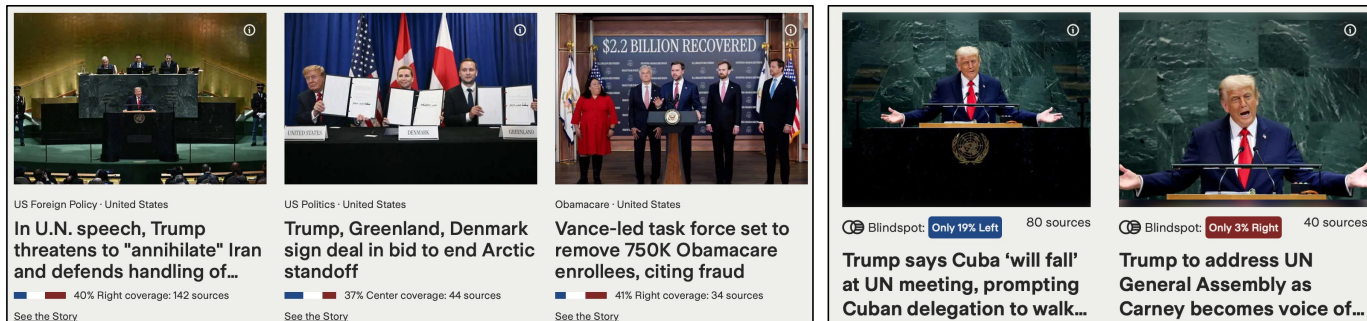


Courthouse News Service

Courthouse News Service – Bias and Credibility Courthouse News Service is rated Least biased with High factual reporting by Media ...

Homework 7 Review

- Media bias and headlines
 - is it predictable from **headline content**?
 - what makes us pick **Left, Center** or **Right**?
 - can we expect even coverage? **No, blindspots**

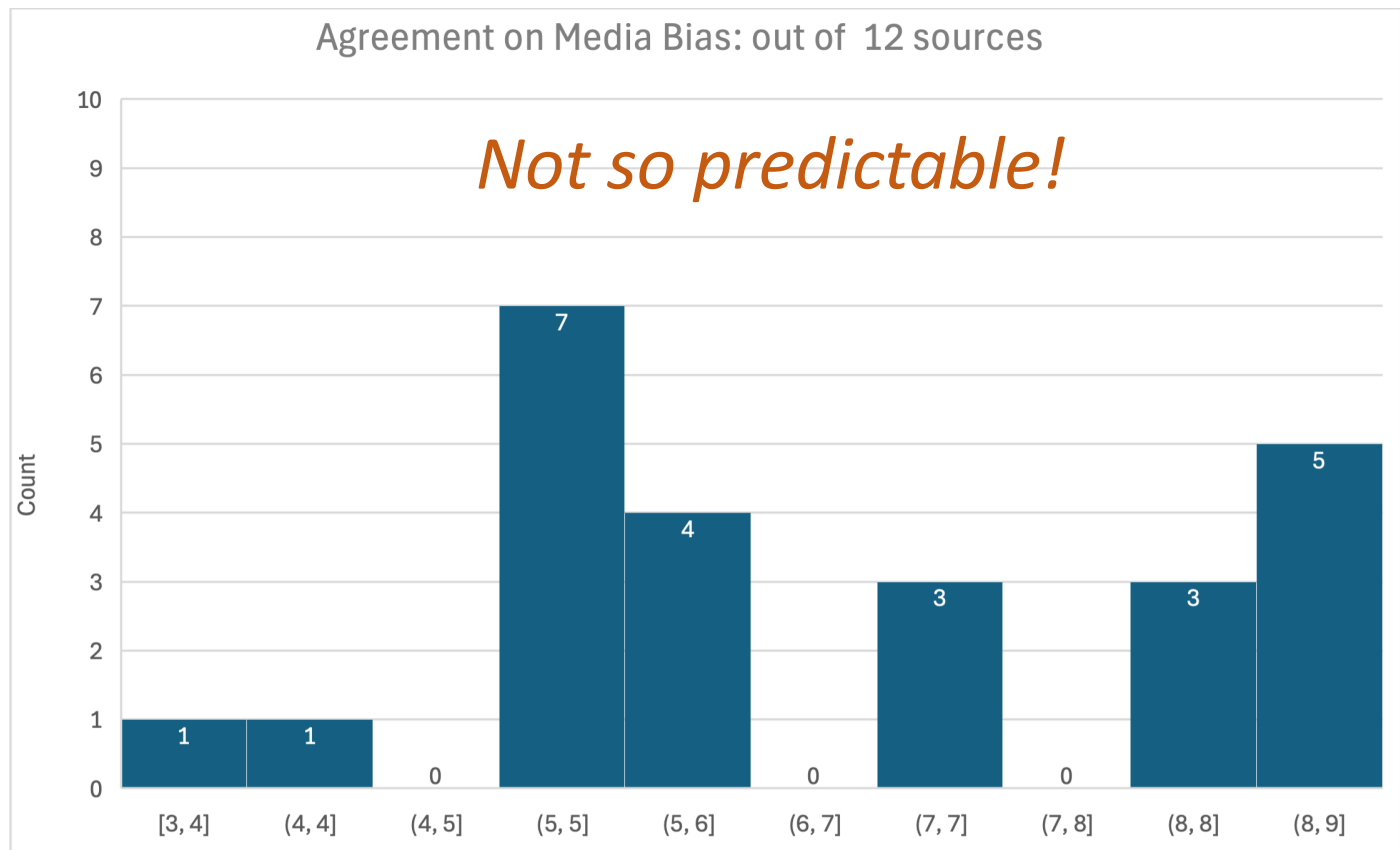


Homework 7 Review

Results:

*strange binning by
Excel btw*

We could do some
modern statistics,
e.g. wrt. inter-
annotator
agreement



Homework 7 Review

How people are labeled:
(political) pejorative terms
for people, groups, etc.
from **Wikipedia**

Pages in category "Pejorative terms for people"

The following 141 pages are in this category, out of 141 total. [This list may not reflect recent changes.](#)

A

- Absent-minded*
- American-born confused desi
- Anchor baby
- Aneran
- Animal epithet
- Antiquarian
- Armchair revolutionary
- Armchair warrior
- Asshole
- Atel (slang)
- Auervaar*

B

- Barbarian
- Basic (slang)
- Beauty beggar
- Benny (slang)
- Black sheep
- Bogan
- Bought priesthood
- Breeder (slang)
- Brigade (pejorative)
- Bro culture
- Buffoon*
- Butthole*

C

- Carpetbagger
- CAVE people*
- Chigyū
- Club fighter
- Coarse (behavior)*
- Contemptuous*
- Control freak
- Coward
- Creonte

D

- Deadbeat parent
- Deviant*
- Disdainful*
- Disgusting*
- Dogmatic*

E

- Edgelord
- Egghead
- Egotistical*
- Emmet (Cornish)
- Envious*

F

- Famous for being famous
- Fantasizer*
- Feral (subculture)
- Fluffy bunny (Wicca)
- FNG syndrome

- Ginge (pejorative)*
- Gold bug

H

- Hack (comedy)
- Hack writer
- Haughty*
- Helicopter parent
- Hipster (contemporary subculture)
- Huan-a

I

- Ignorant*
- Industry plant
- Inept*
- Insignificant*

J

- Jackass (insult)*
- Jobsworth

K

- Knacker

L

- Lace curtain and shanty Irish
- Lamer
- Liar*
- Libertine
- Line stander
- Loan shark
- Loner
- Low-life
- Luser
- Lutyns media

M

- Malinchism
- Mammoni
- Mankurt
- Marites (slang)
- Meddler*
- Mockney
- Moldy figs
- Mundane

N

- Naive*
- Nepo baby
- Nine-headed bird
- Nonsensical*
- Normie (slang)*
- NPC (meme)

O

- Obrazovanshchina
- OK boomer
- Out of touch (phrase)*

P

- Parasite single

- Phillyism*
- Plant epithet
- Pogue
- Poseur
- Prawn sandwich brigade
- Prig
- Pride
- Puddle pirate
- Purteen
- Pussy

R

- Race traitor
- Raja Beta syndrome
- Red neck*
- Reply guy
- Ridiculous
- Rube*
- Running dog

S

- Schilmazel*
- Schnorrer
- Script kiddie
- SheigetZ
- Shiksa
- Shoobie
- Shyster
- Simp
- Slacker
- Snowflake (slang)
- Softy
- Square (slang)
- Sucker (slang)*
- Swine

T

- Tin foil hat
- Tomato can
- Trolling

U

- Un-Australian
- Unconventional*
- Uncool*
- Undesirable*
- Uninteresting*
- Unrefined (behavior)*
- Useful idiot

V

- Vacuous

W

- Walter Mitty
- Wowser

Y

- Yahoo (Gulliver's Travels)
- Yob (slang)

Homework 7 Review

How people are labeled:
(political) pejorative terms
for people, groups, etc.
from **Wikipedia**

Pages in category "Political pejoratives for people"

The following 101 pages are in this category, out of 101 total. [This list may not reflect recent changes.](#)

A

- [ACAB](#)
- [Anglosaksy](#)
- [Anti-national \(India\)](#)
- [Apparatchik](#)
- [Benedict Arnold](#)

B

- [Baizuo](#)
- [Banderite](#)
- [Beltway bandit](#)
- [Béni-oui-oui](#)
- [Bhakt \(pejorative\)](#)
- [BlueAnon](#)
- [Boba liberal](#)
- *[Bobroedka](#)*
- [Bolsominion](#)

C

- [Chairo \(slang\)](#)
- [Champagne socialist](#)
- [Chickenhawk \(politics\)](#)
- [Chud \(pejorative\)](#)
- [Class traitor](#)
- *[Commie](#)*
- [Counter-revolutionary](#)
- [Croppy](#)
- [Cuckservative](#)

D

- [Democrat in name only](#)
- *[Do-gooder](#)*
- [Doughface](#)

E

- [Etrog \(political term\)](#)
- [Eurocrat](#)

F

- [Fascist \(insult\)](#)
- [Fellow traveller](#)
- [Feminazi](#)
- [Flamingant](#)

G

- [Grandpa in his bunker](#)
- [Gukppong](#)
- [Gusano](#)

- [Gutmensch](#)

H

- [Hanjian](#)
- [Homo Sovieticus](#)
- [House slave](#)

K

- [Kadrun](#)

L

- [Liberal elite](#)
- [Limousine liberal](#)
- [Little Eichmanns](#)
- [Little Pink](#)
- [Loony left](#)
- [Low information voter](#)
- [Luddites](#)

M

- [Mayberry Machiavelli](#)
- [Moonbat](#)
- [Morenazi](#)

N

- *[Nazi](#)*
- [List of nicknames used by Donald Trump](#)
- [NIMBY](#)
- [Nomenklatura](#)

O

- [Obama is a schmuck](#)

P

- [Social parasitism \(offense\)](#)
- [Peñabot](#)
- [Performative male](#)
- [Pinko](#)
- [Podkulachnik](#)
- [Political hack](#)
- [Putin khuylo!](#)
- [Putinversteher](#)
- [Putler](#)

Q

- [Quisling](#)

R

- [Radical chic](#)

- [Razakars \(Pakistan\)](#)
- [Reactionary](#)
- [Republican in name only](#)
- [Republicrat](#)
- [Rootless cosmopolitan](#)

S

- [Salim \(Thai slang\)](#)
- [Scalawag](#)
- [Self-hating Jew](#)
- *[Sheeple](#)*
- [Snowflake \(slang\)](#)
- [Social justice warrior](#)
- [Sovok](#)
- [Spiritually Israeli](#)
- [Spiritually Japanese](#)
- [Strikebreaker](#)

T

- *[TACO](#)*
- [Tankie](#)
- [TERF \(acronym\)](#)
- [Terruqueo](#)
- [Tick \(pejorative\)](#)
- [Titushky](#)
- [Tochak Waegu](#)
- *[Totalitarian](#)*
- [Two-faced person](#)
- [Tyrant](#)

U

- [Uncle Tom](#)
- [Upper-class citizen](#)
- [Useful idiot](#)

V

- [Vatnik](#)
- [Viết gлян](#)
- [Volksschädling](#)

W

- [Woman card](#)

X

- [Xiaoyao Faction](#)

Z

- [Zionist as a pejorative](#)
- [Zydokomuna](#)

Homework 7 Review

How people are labeled:
(political) pejorative terms
for people, groups, etc.
from **Wikipedia**

Pages in category "Political pejoratives"

The following 97 pages are in this category, out of 97 total. [This list may not reflect recent changes.](#)

0–9

- 50 Cent Party
- 1450 Internet army

A

- Asylum shopping

B

- Balkanization
- Beauty trends among American conservatives
- Big government
- Bourgeois nation

C

- Cancel culture
- Chinese virus (politics)
- Class reductionism
- Communist bandit
- Criticism of the Quran
- Crypto-communism
- Crypto-fascism

D

- Death panel
- Democrat Party (epithet)
- Developing country
- Deviant current
- Dhimmitude
- Dummymandering

E

- Eco-terrorism
- Ecofascism
- Economism

F

- F-Haine*
- Fatahland
- Feudal fascism

G

- Gay agenda
- Gayrope
- Global arrogance
- Godi media
- Goose step
- Great Satan

- Green imperialism
- Green Terror (Taiwan)

H

- Hamastan
- Hillarycare*

I

- Immanentize the eschaton
- Imperial Manila
- Islamofascism
- Islamofascism

K

- Kangaroo court

L

- Lebanization
- Lemon socialism
- List of United States political catchphrases
- Little Satan
- Livingstone Formulation
- Lois scélérates
- Loibertarianism*
- Londonistan
- Luxury belief
- Lying press

M

- M-She-U*
- The Man
- Menshevizing idealism
- Merchants of death
- Miami Mafia
- Miscegenation

N

- Nanny state
- Nation of shopkeepers
- Nerus (political slur)
- Netroots
- Noynoying
- Nutroots*

O

- Obamacare*
- Old fox

- Omnicause*
- Overclass

P

- Parliamentary cretinism
- Partitionism
- Peace disease
- Pensée unique
- Perfidious Albion
- PIGS (economics)
- Political correctness
- Pork barrel
- Putinland

R

- R-Haine*
- Reactionary centrism
- Red bull (slang)
- Red-baiting
- Revisionism (Marxism)

S

- Scientism
- Sedition Caucus
- Shooting and crying
- Sick man of Asia
- Sick man of Europe
- Slacktivism
- Social fascism

T

- Tax and spend
- Taxachusetts*
- Third World
- Tory scum
- Trump derangement syndrome

V

- Victim feminism

W

- Wahhabi (epithet)
- Waving the bloody shirt

Z

- Zombie liberalism

Homework 7 Review

- Media bias and headlines
 - How about **AI**?
 - Does it do a better job than **you**?

Tell me if you think this headline feels like it is from a left, center or right wing source? Can you explain why you think so? Just use the words in the headline only for your reasoning.

Homework 7 Review

2 for 2

1. Milo Yiannopoulos says he was 'terrified' by the 'brutality' of his ICE deportation

Using **only the words in the headline**, I would characterize it as **center-left leaning in framing**, though not strongly so.

Why

The strongest signals are:

- **"terrified"** — an emotionally charged description of the person's experience. Including the emotional reaction rather than simply reporting the deportation makes the headline more sympathetic to him.
- **"brutality"** — this is the biggest clue. "Brutality" is a highly negative characterization of ICE's conduct. The quotation marks make clear that the headline is **attributing** the characterization to Yiannopoulos rather than stating it as fact.
- **"ICE deportation"** — relatively neutral terminology by itself.

1. Shed no tears for Milo Yiannopoulos, the deported right-wing provocateur

Based **only on the words in the headline**, I'd say **left-leaning**.

Why:

- **"Shed no tears"** is openly dismissive. It tells the reader what emotional reaction they *shouldn't* have, rather than neutrally describing what happened.
- **"provocateur"** is a loaded characterization. It frames Yiannopoulos primarily as someone who deliberately provokes controversy rather than simply identifying his political role.
- **"right-wing"** is descriptive in itself, but paired with "provocateur" and "shed no tears," it contributes to a distinctly negative framing.
- **"deported"** is relatively neutral and factual.

The strongest signal is **"Shed no tears"**: a center headline would be more likely to say something like *"Milo Yiannopoulos, right-wing provocateur, deported"* without prescribing the reader's emotional response.

So, from the headline alone: **Left** — primarily because of its openly dismissive and evaluative language.

Homework 7 Review

2 for 4

1. I am a persecuted human rights icon

Left-leaning, or possibly center-left.

Why, based **only on the words in the headline**:

- **"persecuted"** is emotionally charged and frames the subject as a victim of injustice.
- **"human rights"** is strongly associated with progressive/left-of-center political vocabulary, although it is not exclusively left-wing.
- **"icon"** is highly laudatory—it elevates the person rather than neutrally describing them.
- The first-person **"I am"** gives it a dramatic, personal, self-identifying tone.

Overall, the combination of **"persecuted"** + **"human rights"** + **"icon"** makes it feel more **left/center-left** than center or right.

1. Milo Yiannopoulos, cheerleader of mass deportations, detained by ICE

Feels left-wing.

Why, using only the headline's words:

- **"cheerleader"** is a loaded, dismissive characterization rather than a neutral description.
- **"mass deportations"** foregrounds a controversial policy in stark terms.
- The structure **"cheerleader of mass deportations"** frames Milo Yiannopoulos through criticism of his political position rather than simply identifying him.
- **"detained by ICE"** is relatively neutral, but combined with the preceding phrase, the overall framing feels adversarial toward him.

So, based **strictly on the headline's wording**, I'd classify it as **left-wing**.

Homework 7 Review

3 for 6

1. Deported Milo Yiannopoulos says he is 'ashamed' of support for Trump

Feels left-leaning.

Using **only the headline's words**:

- **"Deported"** — presents Milo Yiannopoulos through a negative or consequential event.
- **"ashamed"** — strongly negative emotional language.
- **"support for Trump"** — explicitly connects him to Trump, while the word "ashamed" frames that association negatively.

The headline doesn't use an overtly insulting label, but the combination of **"deported"** + **"ashamed"** + **"support for Trump"** gives it a distinctly negative framing toward a Trump supporter.

1. Milo Yiannopoulos deported to the U.K. one day after ICE arrest

1. Center

Based **only on the wording of the headline**, it feels relatively **center/straight-news**.

Why:

- **"deported"** is a neutral, factual description.
- **"to the U.K."** simply states the destination.
- **"one day after ICE arrest"** gives a specific chronology rather than an emotional interpretation.
- There are **no loaded adjectives or ideological labels** such as "far-right," "provocateur," "victim," "illegal," or "crackdown."

The headline is notably **spare and procedural**. It tells you *what happened and when*, without signaling approval or disapproval.

Homework 7 Review

5 for 8

1. Milo Yiannopoulos' Black husband accuses him of shocking racism, "enslavement" & abuse

Left-leaning.

Why, based only on the headline:

- **"shocking racism"** is strongly condemnatory language rather than neutral description.
- **"enslavement"** is an emotionally charged term that presents a severe allegation.
- **"abuse"** adds another strongly negative characterization.
- **"accuses him"** clearly foregrounds allegations against Yiannopoulos.
- **"Black husband"** foregrounds race, which contributes to the headline's framing.

Overall, the combination of **"shocking," "racism," "enslavement," and "abuse"** gives the headline a strongly critical framing that feels more **left-wing than center or right-wing**.

1. Milo Yiannopoulos 'very ashamed' of support for Trump after deportation

Feels left-leaning.

Using only the headline's words:

- **"very ashamed"** is a strongly negative characterization of the subject's past political support.
- **"support for Trump"** explicitly identifies Trump as the object of that regret.
- **"after deportation"** creates a cause-and-effect narrative between the deportation and the change in attitude.

The wording is not neutral: it foregrounds **regret and repudiation of Trump**, which gives it a distinctly left-leaning feel.

Homework 7 Review

6 for 10

1. *Milo Yiannopoulos, Right-Wing Agitator, Deported Home to U.K.*

Left-leaning

Why, based only on the headline's wording:

- **"Right-Wing Agitator"** is a strongly characterizing description rather than a neutral occupational or political label.
- **"Agitator"** has a negative, confrontational connotation and suggests someone who deliberately provokes unrest or controversy.
- **"Deported Home to U.K."** is relatively matter-of-fact, although "home" emphasizes his foreign origin.

The key signal is **"Right-Wing Agitator"**: that phrasing feels more characteristic of a left-leaning source than a center or right-leaning one.

1. Far-right commentator Milo Yiannopoulos deported to UK

Left-leaning, based on the headline wording alone.

- **"Far-right"** is an explicit ideological label. Its use makes the headline feel more characteristic of a left-leaning or progressive publication than a right-leaning one.
- **"Commentator"** is relatively neutral.
- **"Deported to UK"** is straightforwardly factual and doesn't itself signal a political orientation.

So, judging **only the words in the headline**, I'd classify it **left**, with the main signal being **"far-right."**

Homework 7 Review

8 for 12

1. British far-right provocateur Milo Yiannopoulos deported after ICE arrest

Left-leaning, based solely on the wording.

Why:

- **"far-right"** explicitly places Yiannopoulos at the extreme end of the political spectrum rather than using a more neutral description such as "conservative" or "political commentator."
- **"provocateur"** is a negatively charged characterization, suggesting someone who deliberately causes controversy.
- **"deported after ICE arrest"** is relatively factual and neutral, but the first two descriptors give the headline its overall tone.

So, judging **only the headline's language**, I'd classify it as **left**. The wording is noticeably more editorialized than a center-style headline such as "Milo Yiannopoulos deported after ICE arrest."

1. Far-right British provocateur Milo Yiannopoulos arrested by US immigration enforcement

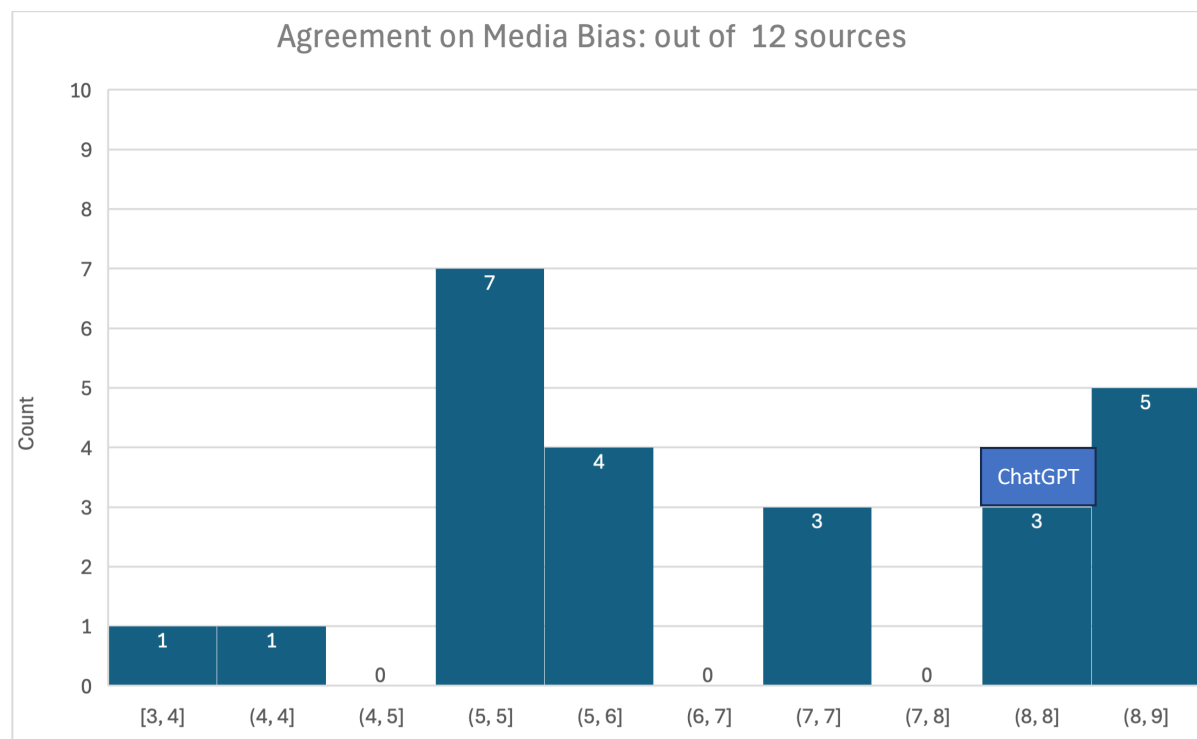
My read: Left-leaning

Why, based only on the headline's wording:

- **"Far-right"** is an explicitly ideological characterization rather than a neutral description.
- **"Provocateur"** is a strongly loaded term, suggesting someone who deliberately creates controversy or agitation.
- **"Arrested by US immigration enforcement"** is relatively factual and neutral compared with the first two descriptors.

The combination of **"far-right"** + **"provocateur"** gives the headline a distinctly critical framing, which makes it feel more **left-leaning** than center or right.

Homework 7 Review



Homework 7 Review

	A	B	C	D
1	Headline AllSides Number Name			
2	1 L		79%	CNN
3	2 L		92%	LA Times
4	3 R		79%	Spectator
5	4 C		29%	Courthouse News
6	5 L		67%	Guardian
7	6 L		0%	NBC News
8	7 L		83%	LGBTQ Nation
9	8 C		29%	The Hill
10	9 L		58%	New York Times
11	10 C		75%	Reuters
12	11 L		29%	ABC News
13	12 L		29%	AP News
14	Agreement		54%	6.52

Interesting read

- <https://slate.com/technology/2019/10/consequential-computer-code-software-history.html>

future tense

The Lines of Code That Changed Everything

Apollo 11, the JPEG, the first pop-up ad, and 33 other bits of software that have transformed our world.

Hello, World!

Date: 1972 or earlier

The phrase that has introduced generations to code

```
main( ) { printf("hello, world\n"); }
```

When you sit down to learn a new programming language, the first thing the tutorial has you do is get the computer to display the phrase “Hello, world!” Perhaps the most famous early example comes from a Bell Laboratories memorandum called “Programming in C—A Tutorial,” written in 1974, though it was also found in a 1972 manual for another language, B, and may go back even earlier than that.

Perl code and History

The Code That Made a T-Shirt Illegal

Date: Circa 1995

Language: Perl

One of the earliest examples of code as activism

```
#!/bin/perl -s-- -export-a-crypto-system-sig -RSA-3-lines-PERL
$m=unpack(H,$w,$m,"\0"x$w),$_=`echo "16do$w 2+40i0$d*-^1[d2%Sa
2/d0<X+d*Lal=z\U$n%0]SX$k"[$m*]\EszlXx++p|dc`,s/^.\|W//g,print
pack('H*',$_)while read(STDIN,$m,($w=2*$d-1+length($n)&~1)/2)
```

[Munitions T-Shirt Homepage](#)

“WARNING: This shirt is classified as a munition and may not be exported from the United States, or shown to a foreign national,” the shirt warned. For a time, the United States government treated strong encryption like surface-to-air missiles: too dangerous to fall into the hands of America’s foes. The idea made a kind of sense when encryption lived in heavy, expensive devices, but a lot less sense when the State Department tried to tell cryptography researchers in the 1990s they couldn’t post their code on the internet. But the RSA encryption algorithm—one of the basic building blocks of modern cryptography—is elegant enough that it can be written out in just four dense lines of Perl code ... short enough to fit on a T-shirt. The original shirts are now collector’s items; the export controls, although not completely gone, have been substantially pared back. —James Grimmelmann, professor of law at Cornell Tech and Cornell Law School

Perl code and History

<http://www.cypherspace.org/adam/rsa/rsa-details.html>

```
#!/bin/perl -sp0777i<X+d*lMLa^*lN%0]dsXx++lMlN/dsM0<j]dsj  
$/=unpack('H*',$_);$._=`echo 16dio\U$k"SK$/SM$n\EsN0p[lN*1  
lK[d2%Sa2/d0$^Ixp"|dc`;s/\W//g;$._=pack('H*',/((..)*$)/)
```



Perl regex

- `\b` or `\b{wb}`

`\b{wb}`

This matches a Unicode "Word Boundary", but tailored to Perl expectations. This gives better (though not perfect) results for natural language processing than plain `\b` (without braces) does. For example, it understands that apostrophes can be in the middle of words and that parentheses aren't (see the examples below). More details are at <http://www.unicode.org/reports/tr29/>.

- Perl global variables set during regex matching:

<code>\$^</code>	Everything prior to matched string
<code>\$&</code>	Entire matched string
<code>\$'</code>	Everything after to matched string

- other boundary metacharacters: `^` (beginning of line), `$` (end of line)

Perl regex: Unicode and \b

\b

```
1 $s = "This isn't a U.A.-approved sentence.";
2 while( $s =~ m/\b(\w.*?)\b/g ) {
3     print "$&\n";
4 }
```

This
isn
t
a
U
A
approved
sentence

Note: global
match in
while-loop

\b{wb}

```
1 $s = "This isn't a U.A.-approved sentence.";
2 while( $s =~ m/\b{wb}(\w.*?)\b{wb}/g ) {
3     print "$&\n";
4 }
```

This
isn't
a
U.A
approved
sentence

Note: `. * ?` is the non-greedy version of `. *`

Perl regex: Unicode and \w

- \w is [0-9A-Za-z_]

Definition is expanded for Unicode:

```
use utf8;
```

```
use open qw(:std :utf8);
```

```
my $str = "school école École šola trường स्कूल škole โรงเรียน";
```

```
@words = ($str =~ /(\w+)/g);
```

```
foreach $word (@words) { print "$word\n" }
```

use **pragma**: <https://perldoc.perl.org/open.html>

list
context

```
bash-3.2$ perl regex_utf.perl
school
école
École
šola
trường
स्कूल
škole
โรงเรียน
```

```
school
cole
cole
ola
tr
ng
kole
```

Perl regex: Unicode and \w

- \w is [0-9A-Za-z_]

Definition is expanded for Unicode:

```
use utf8;
```

```
use open qw(:std :utf8);
```

```
my $str = "school école École šola trường स्कूल škole โรงเรียน";
```

```
@words = ($str =~ /(\w+)/g);
```

```
foreach $word (@words) { print "$word\n" }
```

use **pragma**: <https://perldoc.perl.org/open.html>

list
context

```
bash-3.2$ perl regex_utf.perl
school
école
École
šola
trường
स्कूल
škole
โรงเรียน
```

```
school
cole
cole
ola
tr
ng
kole
```

Chapter 2: JM

RE	Match	Example Patterns Matched
*	an asterisk “*”	“K*_A*_P*_L*_A*_N”
\.	a period “.”	“Dr. Livingston, I presume”
\?	a question mark	“Why don’t they come and lend a hand_?”
\n	a newline	
\t	a tab	

Figure 2.8 Some characters that need to be backslashed.

Why is a backslash needed?

- * means zero or more repetitions of the previous char/expr
- . means any single character
- ? means previous char/expr is optional (zero or one occurrence)

Chapter 2: JM

- Precedence of operators

- Example: Column 1 Column 2 Column 3 ...

- `/Column [0-9]+ */`

- `/(Column [0-9]+ *)*/`

space

- `/house(cat(s|)|)/` (| = disjunction; ? = optional)

- Perl:

- the regex matched by within the pair of parentheses (...) is stored/**captured** in global variables \$1 (and \$2 and so on).

- (?: ...) group but exclude from \$n variable storage

- Precedence Hierarchy:

Parenthesis	()
Counters	* + ? {}
Sequences and anchors	the ^my end\$
Disjunction	

Perl regex

- Recall earlier lecture about time?

<http://perldoc.perl.org/perlretut.html>

```
1. # extract hours, minutes, seconds
2. if ($time =~ /(\d\d):(\d\d):(\d\d)/) { # match hh:mm:ss format
3.     $hours = $1;
4.     $minutes = $2;
5.     $seconds = $3;
6. }
```

returns 1 (true) or "" (empty if false)

A shortcut: **list** context for matching

```
1. # extract hours, minutes, seconds
2. ($hours, $minutes, $second) = ($time =~ /(\d\d):(\d\d):(\d\d)/);
```

returns a list

Chapter 2: JM

Backreferences

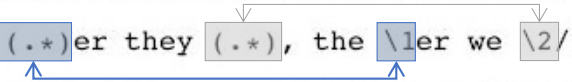
Closely associated with the matching variables `$1`, `$2`, ... are the *backreferences* `\1`, `\2`, ... Backreferences are simply matching variables that can be used *inside* a regexp. This is a really nice feature; what matches later in a

- `s/([0-9]+)/<\1>/`  what does this do?

Backreferences give Perl regexs more expressive power than **Finite State Automata (FSA)**

The number operator can be used with other numbers. If you match two different sets of parenthesis, `\2` means whatever matched the *second* set. For example,

/the `(.*)`er they `(.*)`, the `\1`er we `\2`/



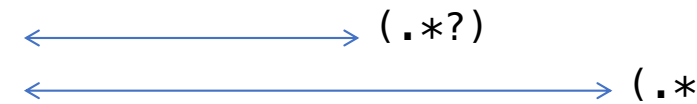
will match *The faster they ran, the faster we ran* but not *The faster they ran, the faster we ate*. These numbered memories are called **registers** (e.g., register 1, register 2,

Shortest vs. Greedy Matching

- default behavior
 - in Perl regex matching:
 - *take the longest possible matching string*
 - *and see if it works*
 - *if so, ok*
 - *if not, **backtrack** (take a shorter match and try again)*
 - aka **greedy matching**
 - *This behavior can be changed, see next slide*

Shortest vs. Greedy Matching

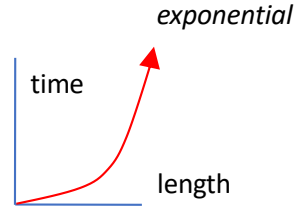
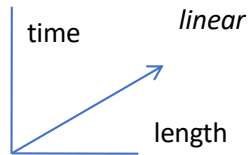
from <http://www.perl.com/doc/manual/html/pod/perlre.html>

- Example:


```
$_ = "The food is under the bar in the barn.";
if ( /foo(.*)bar/ ) {
    print "matched <$1>\n";
}
```
- Output:
 - greedy (.*):
matched <d is under the bar in the >
 - shortest (.*)?:
matched <d is under the >
- Notes:
 - ? immediately following a repetition operator like * (or +) makes the operator work in non-greedy mode
 - + immediately following a repetition operator makes it **non-backtracking** greedy

Regex: exponential time

- Regex search is supposed to be fast
 - but searching is not necessarily proportional to the length of the string (or corpus) being searched
 - in fact, Perl RE matching can take exponential time (in length)



- **non-deterministic**

- *may need to backtrack (revisit last choice point) if it matches incorrectly part of the way through*
- Let's consider `a?a?a?aaa` matching against the string *aaa*

Regex: exponential time

- Consider $a?a?a?aaa$ matching against the string aaa
 - For expository purposes: $a_1 a_2 a_3$
 - red a = failure to match (causes backtracking)

Tries:

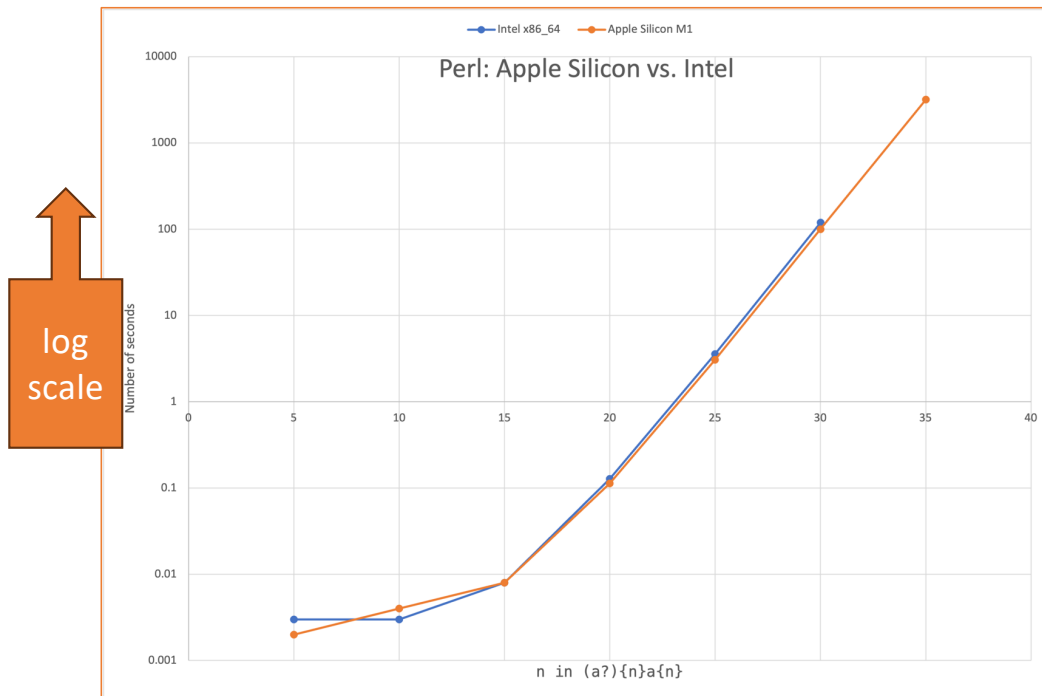
- 
1. $a_1? a_2? a_3? \textcolor{red}{a}aa$
 2. $a_1? a_2? \quad ? a_3 \textcolor{red}{a}a$
 3. $a_1? \quad ? a_2 \quad ? a_3 \textcolor{red}{a}a$
 4. $a_1? \quad ? \quad ? a_2 a_3 \textcolor{red}{a}$
 5. $\quad ? a_1 ? a_2 ? a_3 \textcolor{red}{a}a$
 6. $\quad ? a_1 ? \quad ? a_2 a_3 \textcolor{red}{a}$
 7. $\quad ? \quad ? a_1 ? a_2 a_3 \textcolor{red}{a}$
 8. $\quad ? \quad ? \quad ? a_1 a_2 a_3$



success!

Perl implementations

- Now consider scaling up $a?a?a?aaa$, i.e. $(a?)^na^n$ matching against a^n



- perl 5, version 28, subversion 3 (v5.28.3) built for darwin-thread-multi-2level
 - `/opt/local/bin/perl: Mach-O 64-bit executable x86_64`
- perl 5, version 34, subversion 1 (v5.34.1) built for darwin-thread-multi-2level
 - `/opt/local/bin/perl: Mach-O 64-bit executable arm64`

Reference:

<https://swtch.com/~rsc/regexp/regexp1.html>

Regex: exponential time

- `regex(a?)nan` matching against `an` for a range of values for `n`

```
time perl -e '$n = shift; $na = "a" x $n; print $na =~ /(a?){$n}a{$n}/' 25  
real 0m3.201s  
user 0m3.190s  
sys 0m0.007s
```

- Note:

- `shift` defaults to working on `@ARGV`, that's how `$n` gets 25 above.

shift ARRAY

shift

Shifts the first value of the array off and returns it, shortening the array by 1 and moving everything down. If there are no elements in the array, returns the undefined value. If ARRAY is omitted, shifts the `@_` array within the lexical scope of subroutines and formats, and the `@ARGV` array outside a subroutine and also within the lexical scopes established by the `eval STRING`, `BEGIN {}`, `INIT`