

Morphology, Finite State Machines, and Regular Expressions

CS114 Lab 2
January 26, 2018
Kenneth Lai

Morphology

- Morphology—how to build words out of morphemes
- Morphemes—building blocks for words
 - Stems—“main” morpheme, “main” meaning
 - Affixes—modify the stem

Morphology

- Stems vs. affixes
 - Cat -s
 - Un- believe -able -ly
 - Dog

Morphology

- Inflectional morphology—word similar to stem
 - Same class (part of speech)
 - Similar meaning
- Derivational morphology—word not as similar to stem
 - Different class (part of speech)
 - Not as similar meaning

Morphology

- Inflectional morphology
 - Compute + ing = Computing
- Derivational morphology
 - Compute + er = Computer


Morphology

- Morphological parsing (HW problem 3)
 - Computing = Compute + ing = Compute + present participle
- Stemming
 - Computing = Compute



Uses of Morphology

- Information Retrieval
 - Example from book: Foxes = Fox

Uses of Morphology



fox




All Images News Videos Shopping More Settings Tools

About 1,080,000,000 results (0.64 seconds)


[FOX Broadcasting | TV Shows on FOX | fox.com](#)
[www.fox.com/](#)
See What's On Tonight, Watch Full Episodes & Keep Up With Your Favorite Shows!

Top stories




DOJ recovers missing text messages between anti-Trump FBI agents Strzok and Page

Fox News
1 day ago



Nikki Haley skewers Mahmoud Abbas at UN Security Council meeting

Fox News
2 hours ago



Fox News' Chris Wallace Knocks Talk of 'Deep State Coups' Against Trump: We're...

Mediate
26 mins ago

[More for fox](#)



Stream and Watch your Favorite TV Shows, Movies and Live TV - FOX
[https://www.fox.com/](#)
Watch your favorite FOX TV shows, movies, clips and get all the latest information and scheduling.

Fox News (@FoxNews) · Twitter
[https://twitter.com/FoxNews](#)


.@kingulfoyle: "POTUS has actually done quite a bit here...You've got a housing rebound. The economy's accelerating. Jobs numbers up, infrastructure & development up. There's a lot of positive things happening

.@greggutfeld: "Trump crashes Davos and tells them what's what
[pic.twitter.com/iIRwf0e...](#)

.@JesseBWatters: "It's going to be very difficult for the Democrats to run against winning. I don't see how they do it. How can you deny your 401K exploding? How can you deny tax cuts..."
#TheFive




Media company


[foxnews.com](#)


Fox News is an American basic cable and satellite television news channel owned by the Fox Entertainment Group, a subsidiary of 21st Century Fox. The channel broadcasts primarily from studios at 1211 Avenue of the Americas, New York City, New York. [Wikipedia](#)


Parent organizations: Fox Entertainment Group, 21st Century Fox S&P
Global Market Intelligence


Profiles

[Twitter](#)


[Facebook](#)

[YouTube](#)


[Instagram](#)

[Google+](#)


TV shows




The Five
Since 2011




Fox & Friends
Since 1998



Tucker Carlson Tonight
Since 2016



Special Report
Since 1998




The Story with Martha MacCallum
Since 2017

[View 45+ more](#)



Topics they write about

Crime

Police



foxes



All Images Videos Shopping News More Settings Tools

About 1,080,000,000 results (0.62 seconds)

[Fox - Wikipedia](#)
[https://en.wikipedia.org/wiki/Fox](#)
Foxes are small-to-medium-sized, omnivorous mammals belonging to several genera of the family Canidae. Foxes have a flattened skull, upright triangular ears, a pointed, slightly upturned snout, and a long bushy tail (or brush). Twelve species belong to the monophyletic group of Vulpes genus of "true foxes".
[Foxes](#) · [Foxes in popular culture](#) · [Red fox](#) · [Canidae](#)

People also ask

What is the plural form of the word Fox?

Are dogs and foxes related?

What does a fox do?

What kind of noise does a fox make?

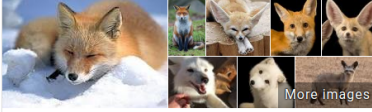
[Feedback](#)

Foxes: Facts & Pictures - Live Science
[https://www.livescience.com/Animals](#)
Sep 14, 2017 - Foxes are omnivorous mammals that are light on their feet. They are often mistaken for other members of the Canidae family, which include jackals, wolves and dogs. They stand out from their relatives because of their long, thin legs, lithe frame, pointed nose and bushy tail. These animals are very social ...

Foxes - Home | Facebook
[https://www.facebook.com/iamfoxes/](#)
Foxes. 1317728 likes · 942 talking about this. Get my new album All I Need here:
[http://iamfox.es/AllINeedTDLX](#).

14 Fascinating Facts About Foxes | Mental Floss
[mentalfloss.com/article/59739/14-fascinating-facts-about-foxes](#)
Sep 23, 2017 - Foxes are part of the Canidae family, which means they're related to wolves, jackals, and dogs. They're medium-sized, between 7 and 15 pounds, with pointy faces, lithe frames, and bushy tails. But unlike their relatives, foxes are not pack animals. When raising their young, they live in small families —called ...

Learn about foxes | Mass.gov
[https://www.mass.gov/service-details/learn-about-foxes](#)




[More images](#)

Fox
Animal


Foxes are small-to-medium-sized, omnivorous mammals belonging to several genera of the family Canidae. Foxes have a flattened skull, upright triangular ears, a pointed, slightly upturned snout, and a long bushy tail.
[Wikipedia](#)

Lifespan: Red fox: 2 – 5 years
Speed: Red fox: 31 mph
Mass: Fennec fox: 1.5 – 3.5 lbs, Red fox: 4.9 – 31 lbs, Arctic fox: 7.1 – 21 lbs
Height: Fennec fox: 8 in., Red fox: 14 – 20 in., Arctic fox: 9.8 – 12 in.
Gestation period: Fennec fox: 50 – 52 days, Red fox: 49 – 58 days, Arctic fox: 52 days
Japanese word: Kitsune


Representative species [View 20+ more](#)




Fennec fox




Red fox



Arctic fox




Culpeo



Gray fox

[Feedback](#)

See results about

Foxes (Singer-songwriter)
Born: April 29, 1989 (age 28), Southampton, United Kingd...
Songs: Youth, Let Go for Tonight, Holding Onto Heaven, D...


Uses of Morphology

- Machine translation (and natural language generation more generally)
- Loves = love + s = love + 3SG = ama

	Present Indicative	Imperfect Indicative	Future	Preterite	Present Subjunctive	Conditional	Imperfect Subjunctive	Future Subjunctive
1SG	amo	amaba	amaré	amé	ame	amaría	amara	amare
2SG	amas	amabas	amarás	amaste	ames	amarías	amaras	amares
3SG	ama	amaba	amará	amó	ame	amaría	amara	amáreme
1PL	amamos	amábamos	amaremos	amamos	amemos	amaríamos	amáramos	amáremos
2PL	amáis	amabais	amaréis	amasteis	améis	amaríais	amarais	amareis
3PL	aman	amaban	amarán	amaron	amen	amarían	amaran	amaren

Figure 3.1 To love in Spanish. Some of the inflected forms of the verb *amar* in European Spanish. *1SG* stands for “first person singular”, *3PL* for “third person plural”, and so on.

Finite State Machines

- Computers with almost no memory
- Useful for morphology (but not syntax)

Finite State Machines

- Automata
 - Output: just accept or reject
- Transducers
 - Output: an output string (or anything you want)

Finite State Machines

- Input: a string (containing symbols from some alphabet)
 - Read one symbol at a time from left to right
- Start in the start (initial) state
- Transition to next state according to current state and read symbol

Finite State Machines

- If you get to the end of the input and you are in:
 - Accept (final) state: return “accept”/output string
 - Other state: return “reject”/do not output string

Finite State Machines

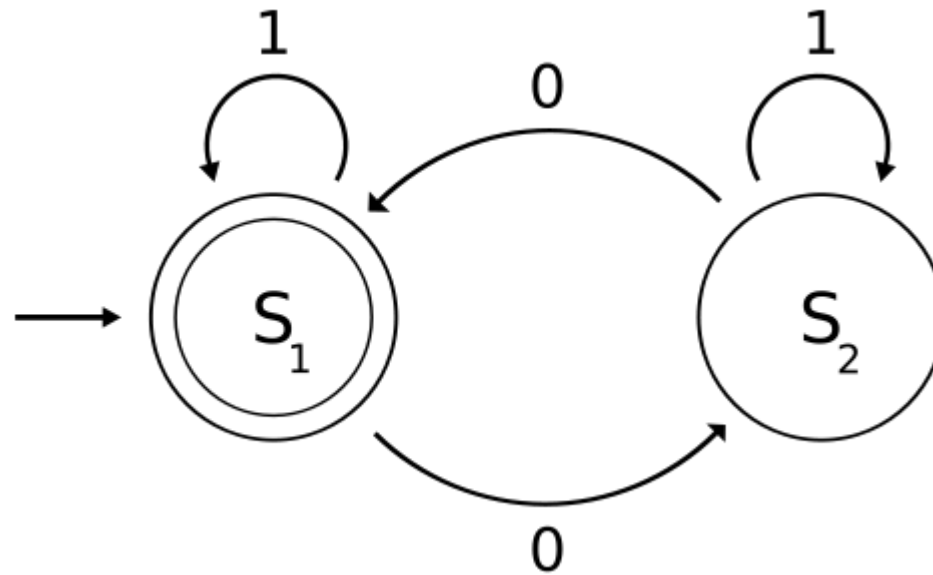
- Determinism—exactly one transition for each combination of state/symbol
- Nondeterminism—zero or more transitions for each combination of state/symbol

Finite State Machines

- For transducers, distinguish between nondeterministic transducers and nondeterministic transductions
 - Transducer—as before
 - Transduction—more than one possible output string for any input string
 - For HW, make sure all transductions are deterministic

Finite State Machines

- State diagram



Finite State Machines

- Language recognized by a finite state machine
 - Set of input strings for which the machine returns “accept”/output string
- A language is regular if and only if it is recognized by a finite state machine

Regular Expressions

- In theory
 - A sequence of characters that defines a regular language
- In practice
 - A sequence of characters that defines a search pattern (not necessarily regular)

Regular Expressions

- Base cases
 - The empty set \emptyset
 - The empty string ε
 - A symbol from the alphabet

Regular Expressions

- Regular operations
 - Union
 - $a|b$
 - Concatenation
 - ab
 - Kleene star
 - a^*

Regular Expressions

- Other things you might see
 - `[A-Za-z]`
 - `^a`
 - `a?`
 - `a+`
 - `\d`
 - `\1`
- Some (but not all) of these are syntactic sugar

Regular Expressions in Python

- NLTK book
- Also: re in the Python Standard Library