

**Proszę nie przewijać tych slajdów myszką.
Zamiast tego proszę je pobrać na dysk,
otworzyć, powiększyć (Ctrl+plus), a potem
nacisnąć klawisz ze strzałką w prawo**

Wyrażenia regularne

dr inż. Marcin Ciura

marcin.ciura@uken.krakow.pl

Wydział Bezpieczeństwa i Informatyki UKEN

Plan na dziś

- Wyrażenia regularne
- Automaty
- Wyrażenia regularne w języku Python
- ...a ponadto wiadomości o dwóch matematykach, kilka zagadek i kilka ciekawostek

Wyrażenia regularne

Co to są wyrażenia regularne?

Wyrażenia regularne to programy, które służą do wyszukiwania wzorców w tekstach

Te programy są zapisane w specjalizowanym języku programowania :-)

Programy w tym języku można zagnieździć w programach, napisanych w innych językach programowania

Kto wymyślił wyrażenia regularne?

Stephen Cole Kleene (5.1.1909–25.1.1994)



Amerykański matematyk, znany z pewnej gwiazdki. Stworzył teorię obliczalności. Wspinął się po górach, działał na rzecz ochrony przyrody. W 1951 roku, kiedy badał sieci neuronowe, wymyślił wyrażenia regularne

Przykład

Każdy gen RNA pasuje do tego wyrażenia regularnego:

(AUG|CUG|UUG)(...)*?(UAA|UAG|UGA)

Regex, regexp

Regex i regexp to skróty angielskiego wyrażenia regular expression
Programiści często używają tych skrótów zamiast pełnej nazwy

Definicje

- **Alfabet**: skończony zbiór **symboli** czyli **znaków**, na przykład {**A, C, G, T**}, ASCII, Unicode...
- **Łańcuch znaków**, krócej **łańcuch**: to samo, co ciąg znaków, na przykład **ATGTGA**
- **Długość łańcucha**: liczba znaków w tym łańcuchu. Oznaczamy długość łańcucha, obejmując ten łańcuch kreskami pionowymi, na przykład $|ATGTGA| = 6$
- **Łańcuch pusty**: łańcuch o długości 0 znaków. Oznaczamy łańcuch pusty grecką literą epsilon: **ϵ**

Definicje

- Wyrażenie regularne to taki łańcuch, który opisuje pewien zbiór łańcuchów zgodnie z pewnymi regułami
Mówimy o każdym łańcuchu z tego zbioru, że ten łańcuch pasuje do danego wyrażenia regularnego
Możemy też powiedzieć, że zbiór łańcuchów opisany przez pewne wyrażenie regularne pasuje do tego wyrażenia regularnego

Elementarne wyrażenia regularne

Każde elementarne wyrażenie regularne składa się z jednego symbolu

- Do symbolu zbioru pustego \emptyset pasuje pusty zbiórłańcuchów: \emptyset
- Do symbolu pustegołańcucha ϵ pasuje zbiór $\{\epsilon\}$, który zawiera tylkołańcuch pusty
- Do każdego wyrażenia regularnegozłożonego z jednegoznaku pasujejedenłańcuch o długości 1. Tenłańcuch składa się z tego samegoznaku, co wyrażenie regularne

Na przykład do wyrażenia regularnegoa pasuje zbiórłańcuchów $\{a\}$:-)

Złożone wyrażenia regularne

- nawiasy
- konkatenacja wyrażeń regularnych
- alternatywa wyrażeń regularnych
- gwiazdka Kleene'a

Złożone wyrażenia regularne: nawiasy

Jeśli R jest wyrażeniem regularnym, to:

- wyrażenie regularne (R) oznacza to samo, co wyrażenie regularne R :-)

Złożone wyrażenia regularne: konkatenacja

Jeśli R i S są wyrażeniami regularnymi, to:

- Do wyrażenia regularnego RS , czyli do konkatenacji wyrażeń regularnych R i S , pasują takiełańcuchy, które powstają, gdy łączymy dowolnyłańcuch pasujący do R z dowolnymłańcuchem pasującym do S

Wyraz konkatenacja pochodzi od angielskiego czasownika *to concatenate*, czyli łączyć włańcuch

Złożone wyrażenia regularne: konkatenacja

Przykłady konkatenacji wyrażeń regularnych:

- Jeśli do R pasuje zbiórłańcuchów $\{d\}$,
a do S pasuje zbiórłańcuchów $\{o\}$,
to do RS pasuje zbiórłańcuchów $\{do\}$
- Jeśli do R pasuje zbiórłańcuchów $\{do, od\}$,
a do S pasuje zbiórłańcuchów $\{dać, pisać\}$,
to do RS pasuje zbiórłańcuchów $\{dodać, dopisać, oddać, odpisać\}$

Złożone wyrażenia regularne: alternatywa

Jeśli R i S są wyrażeniami regularnymi, to:

- Do wyrażenia regularnego $R|S$, czyli do alternatywy wyrażeń regularnych R i S , pasuje suma dwóch zbiorów: zbioru takichłańcuchów, które pasują do R i zbioru takichłańcuchów, które pasują do S

Złożone wyrażenia regularne: alternatywa

Przykłady alternatywy wyrażeń regularnych:

- Jeśli do R pasuje zbiórłańcuchów $\{do\}$,
a do S pasuje zbiórłańcuchów $\{od\}$,
to do $R|S$ pasuje zbiórłańcuchów $\{do, od\}$
- Jeśli do R pasuje zbiórłańcuchów $\{ręka, nogą\}$,
a do S pasuje zbiórłańcuchów $\{ręka, głowa\}$,
to do $R|S$ pasuje zbiórłańcuchów $\{ręka, nogą, głowa\}$

Złożone wyrażenia regularne: gwiazdka Kleene'a

Jeśli R jest wyrażeniem regularnym, to:

- Do wyrażenia regularnego R^* , czyli do domknięcia Kleene'a wyrażenia regularnego R pasuje zbiór takichłańcuchów, które powstają, gdy łączymy zero lub więcejłańcuchów pasujących do R
Domknięcie Kleene'a nazywa się również gwiazdką Kleene'a

Złożone wyrażenia regularne: gwiazdka Kleene'a

Przykłady użycia gwiazdki Kleene'a:

- Jeśli do R pasuje zbiórłańcuchów $\{A\}$, to do R^* pasuje zbiórłańcuchów
 $\{\epsilon, A, AA, AAA, AAAA, AAAAA, AAAAAA, AAAAAAAA, \dots\}$
- Jeśli do R pasuje zbiórłańcuchów $\{fa, sol\}$, to do R^* pasuje zbiórłańcuchów
 $\{\epsilon,$
 $fa, sol,$
 $fafa, fasol, solfa, solsol,$
 $fafafa, fafasol, fasolfa, fasolsol,$
 $solfafa, solfasol, solsolfa, solsolsol,$
 \dots
}

Złożone wyrażenia regularne: kolejność działań

Wykonujemy działania w takiej kolejności:

- najpierw wykonujemy działania w nawiasach
- potem stosujemy gwiazdkę Kleene'a
- potem konkatenujemy wyrażenia regularne
- potem budujemy alternatywy wyrażeń regularnych

Zagadka

Które z dwóch wyrażeń regularnych po prawej stronie znaku = jest równoważne wyrażeniu regularnemu po lewej stronie znaku =?

1. $ab^* = a(b^*)$ czy $(ab)^*$?
2. $a|b^* = a|(b^*)$ czy $(a|b)^*$?
3. $ab|cd = (ab) | (cd)$ czy $a(b|c)d$?

Kolejność działań: nawiasy, gwiazdka, konkatenacja, alternatywa

Zagadka: rozwiążanie

Które z dwóch wyrażeń regularnych po prawej stronie znaku = jest równoważne wyrażeniu regularnemu po lewej stronie znaku =?

1. $ab^* = a(b^*)$
2. $a|b^* = a|(b^*)$
3. $ab|cd = (ab) | (cd)$

Kolejność działań: nawiasy, gwiazdka, konkatenacja, alternatywa

Proszę Państwa o pytania :-)

Automaty

Edward Forrest Moore (23.11.1925–14.6.2003)

Amerykański matematyk i informatyk. W 1956 roku zaproponował konstrukcję takich maszyn, które same będą się rozmnażały. Przypuszczał, że takie maszyny będzie można produkować łatwiej niż statki kosmiczne. W tym samym roku wymyślił taki rodzaj automatu, o jakim będzie mowa na tym wykładzie

Pralka automatyczna



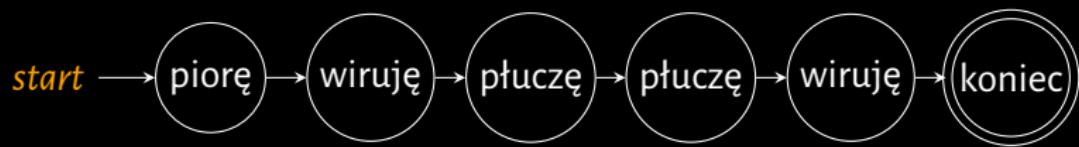
Nie, Edward Forrest Moore nie wynalazł pralki automatycznej :-)

Edward Forrest Moore wymyślił automat skończony

Automat skończony to taki automat, który ma początek i koniec

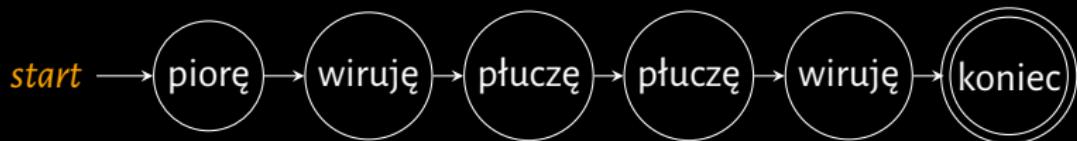
Zrozumiemy automaty skończone dzięki pralce automatycznej

Jak działa pralka automatyczna?



Pralka automatyczna wie, co robi, ale nie pamięta, jak do tego doszła :-)

Jak działa pralka automatyczna?



Kółka oznaczają stany pralki automatycznej

Strzałki oznaczają przejścia między stanami pralki automatycznej

Pralka automatyczna wie, w jakim stanie jest, ale nie pamięta, jak do niego doszła :-)

Automaty skończone

Automaty skończone różnią się od pralki automatycznej tym, że wchodzą znaki

Automaty skończone zmieniają swój stan po każdym wczesnym znaku

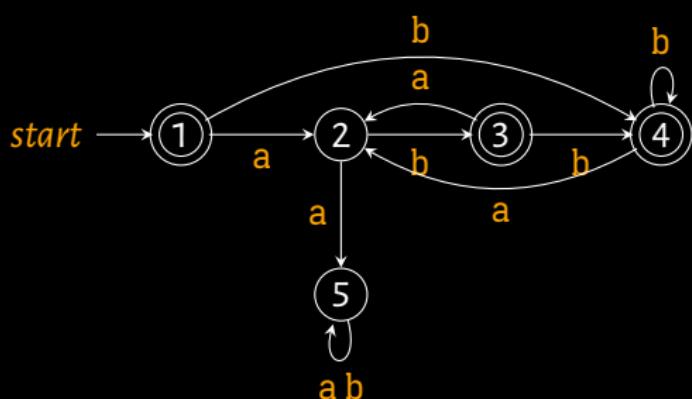
Automat skończony wie, w jakim stanie jest, ale nie pamięta, jak do niego doszedł :-)



Są proste przepisy, czyli algorytmy, które przekształcają wyrażenia regularne na automaty skończone

Tabela przejść i wyjścia automatu skończonego

Tak komputer przechowuje w pamięci automaty skończone:



stan	końcowy?	a	b
1	true	2	4
2	false	5	3
3	true	2	4
4	true	2	4
5	false	5	5

Proszę Państwa o pytania :-)

Wyrażenia regularne w języku Python

Wyrażenia regularne w języku Python

Ta część wykładu jest zapisana w notatniku Jupytera

[Wyrażenia-regularne-w-Pythonie.ipynb](#)

Tę część wykładu opracowałem na podstawie dokumentacji modułu `re`

Dokumentacja tego pakietu znajduje się pod adresem

<https://docs.python.org/3/library/re.html> i na dysku lokalnym

Widzę tę dokumentację na ekranie, gdy w interpreterze Pythona wydaję polecenia

```
import re
```

```
help(re)
```

Proszę Państwa o pytania :-)

Inne zastosowania wyrażeń regularnych

Do tego regexpa pasują daty w formacie ISO 8601, na przykład 1791-05-03:

```
^([0-9][0-9][0-9][0-9])-([0-9]|1[0-2])-([0-2][0-9]|3[01])$
```

Inne zastosowania wyrażeń regularnych

Zwykle używamy przyimka „we” przed takimi wyrazami, które zaczynają się dwiema spółgłoskami, przy czym pierwsza spółgłoska to „f” lub „w”:

we Francji, we Wrocławiu, we foyer

Przed tymi wyrazami też używamy przyimka „we”:

we dwoje, we troje, we czworo

we mnie

we Lwowie, we łbie, we łbach, we łzach, we mgle, we mgłach, we śnie

Inne zastosowania wyrażeń regularnych

Używamy przyimka „we” przed wyrazami, które pasują do tego wyrażenia regularnego:

```
^([fvw]([bcdfghjklłmnprstwżż]|o[iy])  
|dwoje$|troje$|czworo$|mnie$  
|lwow|łb|łz|męg[łł]|śn)
```

Zalecana literatura

<https://kobietydokodu.pl/4-wyrazenia-regularne/>

Dziękuję Państwu za uwagę :-)

Slajdy dodatkowe

Deterministyczne automaty skończone

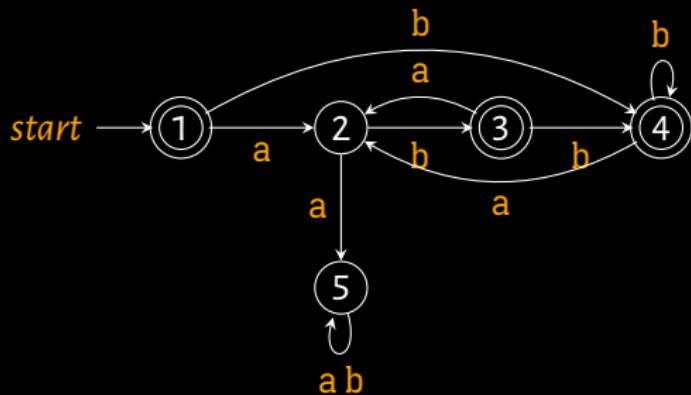
DFA (deterministyczny automat skończony, po angielsku: **Deterministic Finite Automaton**)

Z jednego stanu DFA wychodzi dokładnie jedno przejście oznaczone danym znakiem

Po wczytaniu każdego znaku DFA przechodzi do nowego stanu

Przykład DFA

$$(b^*(a|\varepsilon)b)^*$$



Niedeterministyczne automaty skończone

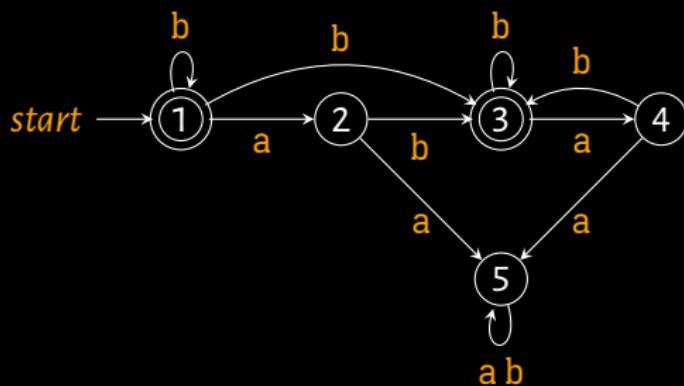
NFA (niedeterministyczny automat skończony, po angielsku:
Nondeterministic Finite Automaton)

Z jednego stanu NFA może wychodzić więcej niż jedno przejście oznaczone tak samo

Po wczytaniu każdego znaku NFA przechodzi do jednego lub więcej stanów naraz

Przykład NFA

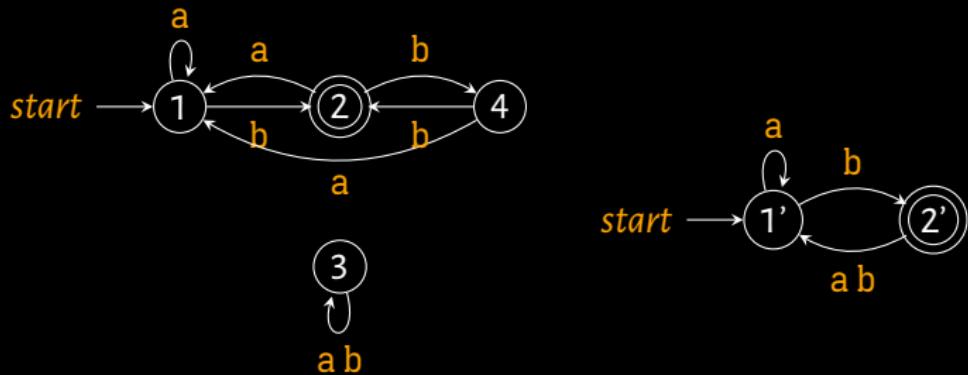
$$(b^*(a|\varepsilon b)^*)$$



Równoważne automaty skończone

Mówimy, że dwa automaty skończone są równoważne, jeśli rozpoznają te same zbiory łańcuchów

Oto przykład dwóch równoważnych automatów skończonych, NFA i DFA:



Źródła zdjęć

https://en.wikipedia.org/wiki/Stephen_Cole_Kleene

<https://my-concept.pl/pl/p/Pralka-automatyczna-SLIM-PP6306s/20820>