Řetězce a seznamy (a kryptografické odbočky)

IB111 Úvod do programování skrze Python

2012

Rozcvička: šifry

1. C S A R B V
   E K T E O A

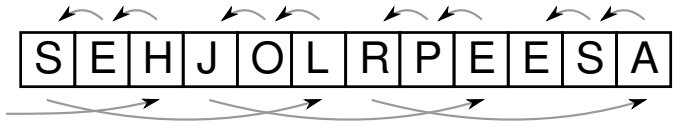
2. C S B U J T M B W B

3. A J L B N O C E

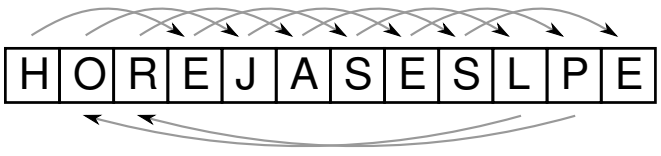
Transpoziční šifry

pozpátku

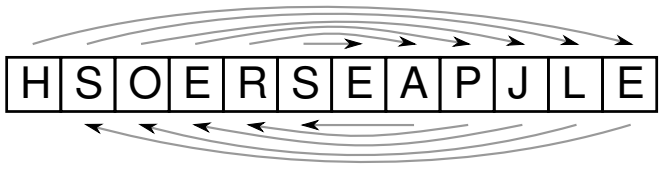
trojice pozpátku



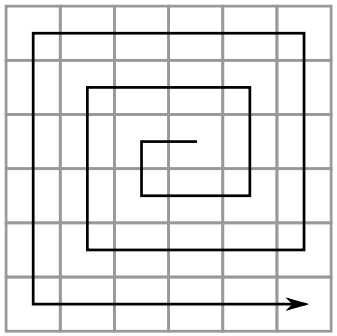
ob tři



dopředu dozadu

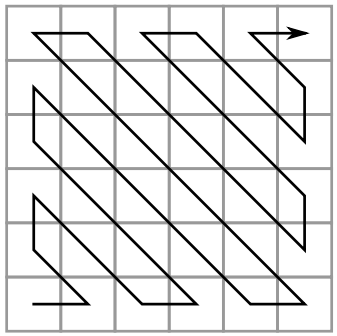


šnek



L	B	A	K	I	N
I	C	S	E	J	B
Z	H	O	P	D	Y
K	O	K	L	A	R
O	V	A	N	Y	U
U	H	R	A	Z	E

cik-cak



N	I	O	U	Z	E
H	B	K	K	H	A
C	O	Y	A	Z	R
L	S	V	R	B	I
K	A	E	A	U	L
P	O	D	J	N	Y

Substituční šifry

Jednoduchá substituce - posun o 3 pozice

	K	O	Z	A
	↓	↓	↓	↓
	10	14	25	0
+3	↓	↓	↓	↓
	13	17	2	3
	↓	↓	↓	↓
	N	R	C	D

Substituce podle hesla

```
HLEDEJPODLIPOU
SLONSLONSLONSL
ZWSQWUDBVWWCGF
```

H → 7

S → 18

7 + 18 → 25 → Z

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
B	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A
C	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
D	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
E	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
F	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E
G	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F
H	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
I	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
J	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
K	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
L	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
M	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
N	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
O	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
P	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Q	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
R	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
S	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
T	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
U	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
V	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
W	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
X	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
Y	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
Z	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y

Řetězce a znaky – ukázky operací

```
"kos" * 3
"petr" + "klic"
text = "velbloud"
len(text)
text[0]
text[2]
text[-1]
ord('b')
chr(99)
```

Řetězce – pokročilejší indexování

```
text = "velbloud"
text[:3]     # první 3 znaky
text[3:]     # od 3 znaku dále
text[1:8:2] # od 2. znaku po 7. krok po 2
text[::3]    # od začátku do konce po 3
```

Řetězce – změny

- neměnitelné (immutable)
- změna znaku – vytvoříme nový řetězec

```
text = "kopec"
text[2] = "n" # chyba
text = text[:2] + "n" + text[3:]
```

Transpozice (rozcvička I)

```
def sifra_po_sloupcich(text,n):
    for i in range(n):
        for j in range(len(text) / n + 1):
            pozice =  j * n + i
            if pozice < len(text):
                print text[pozice],
        print
```

Transpozice (rozcvička I), kratší varianta

```
def sifra_po_sloupcich(text,n):
    for i in range(n):
        print text[i::n]
```

Substituce (rozcvička II)

```
def caesarova_sifra(text, n):
    vystup = ""
    text = text.upper()
    for i in range(len(text)):
        if text[i] == ' ': vystup = vystup + ' '
        else:
            c = ord(text[i]) + n
            if (c > ord('Z')): c = c - 26
            vystup = vystup + chr(c)
    return vystup
```

Caesarova šifra – rozlomení

- máme text zašifrovaný Caesarovou šifrou (s neznámým posunem)
- jak text dešifrujeme?
- příklad: `MPKTWTDVLVELMZCF`

Caesarova šifra – rozlomení

- máme text zašifrovaný Caesarovou šifrou (s neznámým posunem)
- jak text dešifrujeme?
- příklad: `MPKTWTDVLVELMZCF`
- jak to udělat, aby program vrátil jen jednoho kandidáta?

Caesarova šifra – rozlomení

k	Kandidát	b_s	b_f
0	MPKTWTDVLVELMZCF	0	21
1	NQLUXUEWMWFMNADG	13	0
2	ORMVYVFXNXGNOBEH	24	9
3	PSNWZWGYOYHOPCFI	5	-3
4	QTOXAXHZPZIPQDGJ	10	-6
5	RUPYBYIAQAJQREHK	0	9
6	SVQZCZJBRBKRSFIL	0	3
7	TWRADAKCSCLSTGJM	32	26
8	UXSBEBLDTDMTUHKN	0	24
9	VYTCFCMEUENUVILO	11	46
10	WZUDGDNFVFOVWJMP	0	-6
11	XAVEHEOGWGPWXKNQ	5	-2
12	YBWFIFPHXHQXYLOR	0	-28
13	ZCXGJGQIYIRYZMPS	0	-13
14	ADYHKHRJZJSZANQT	0	16
15	BEZILISKAKTABORU	**67**	**59**
16	CFAJMJTLBLUBCPSV	0	11
17	DGBKNKUMCMVCDQTW	5	-4
18	EHCLOLVNDNWDERUX	17	31
19	FIDMPMWOEOXEFSVY	5	22
20	GJENQNXPFPYFGTWZ	4	-23
21	HKFOROYQGQZGHUXA	16	-17
22	ILGPSPZRHRAHIVYB	28	18
23	JMHQTQASISBIJWZC	9	0
24	KNIRURBTJTCJKXAD	5	24
25	LOJSVSCUKUDKLYBE	4	29

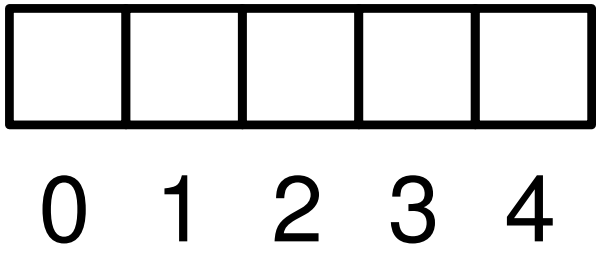
Seznamy (pole) – motivace

- frekvence písmen v textu
- řazení studentů podle bodů na písemce
- reprezentace herního plánu (piškvorky, šachy)

Frekvenční analýza nevhodně

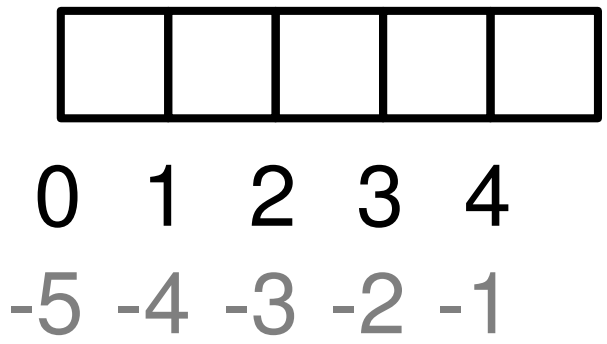
```
def frekvencni_analyza(text):
    frekA = 0
    frekB = 0
    frekC = 0
    for pismeno in text:
        if pismeno == 'A':
            frekA += 1
        elif pismeno == 'B':
            frekB += 1
        elif pismeno == 'C':
            frekC += 1
    print 'A', frekA
    print 'B', frekB
    print 'C', frekC
```

Seznamy (pole)



- „více položek za sebou v pevném pořadí“
- indexováno od nuly!
- základní koncept dostupný ve všech jazycích: „pole“ (array), položky stejného typu, pevně daná délka
- seznamy v Pythonu – obecnější
- Python a pole – knihovna NumPy (nad rámec IB111)

Seznamy v Pythonu



- seznam (list), n-tice (tuple)
- položky mohou být různého typu
- variabilní délka
- indexování i od konce (pomocí záporných čísel)

Seznamy: použití v Pythonu

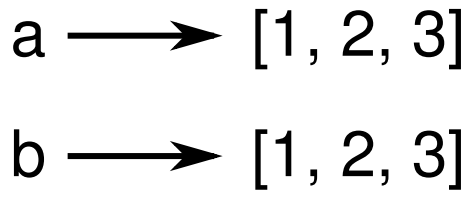
```
s = []          # deklarace prázdného seznamu
s = [3, 4, 1, 8 ]
s[2]            # indexace pruku, s[2] = 1
s[-1]           # indexace od konce, s[-1] = 8
s[2] = 15       # změna pruku
s.append(6)     # přidání pruku
s[1:4]          # indexace intervalu, s[1:4] = [4, 15, 8]
len(s)          # délka seznamu, len(s) = 5
t = [ 3, "pes", [2, 7], -8.3 ]
                # seznam může obsahovat různé typy
```

Python: seznamy a cyklus for

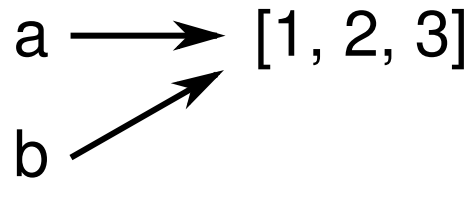
- cyklus `for` – přes prvky seznamu
- `range` – vrací seznam čísel
- typické použití: `for i in range(n):`
- ale můžeme třeba i:
  - `for zvire in ["pes", "kocka", "prase"]: ...`
  - `for pismeno in "velbloud": ...`

Objekty, hodnoty, aliasy

a = [1, 2, 3]
b = [1, 2, 3] nebo b = a[:]



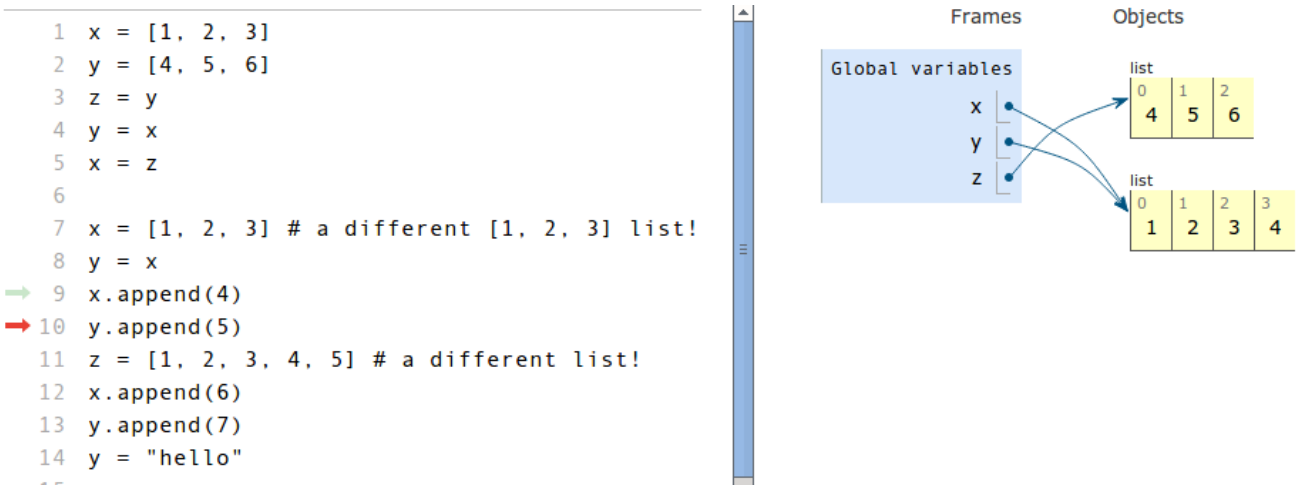
a = [1, 2, 3]
b = a



- parametry funkcí – pouze volání hodnotou
  (na rozdíl např. od Pascalu: volání hodnotou a odkazem)
- měnitelné objekty (např. seznam) však funkce může měnit
- n-tice (tupples) – neměnitelná varianta seznamů
- více na cvičeních, později

Vizualizace běhu programu

`http://www.pythontutor.com/`



vhodné např. pokud je nejasný některý z příkladů ve slidech

Příklad: výpočet průměrné hodnoty

```
def prumer1(seznam):
    soucet = 0.0
    for i in range(len(seznam)):
        soucet += seznam[i]
    return soucet / len(seznam)

def prumer2(seznam):
    soucet = 0.0
    for x in seznam:
        soucet += x
    return soucet / len(seznam)

def prumer3(seznam):
    return float(sum(seznam)) / len(seznam)
```

Ilustrace práce se seznamem

```
def seznam_delitelu(n):
    delitele = []
    for i in range(1, n+1):
        if n % i == 0:
            delitele.append(i)
    return delitele

delitele24 = seznam_delitelu(24)
print delitele24
print len(delitele24)
for x in delitele24: print x**2,
```

Frekvenční analýza lépe

```
def frekvencni_analyza(text):
    frekvence = [ 0 for i in range(26) ]
    for pismeno in text:
        if ord(pismeno) >= ord('A') and\
           ord(pismeno) <= ord('Z'):
            frekvence[ord(pismeno) - ord('A')] += 1
    for i in range(26):
        if frekvence[i] != 0:
            print chr(ord('A')+i), frekvence[i]
```

Převod do Morseovy abecedy

- vstup: řetězec
- výstup: zápis v Morseově abecedě
- příklad: `PES` $\rightarrow$ `.--.|.|...`

Převod do Morseovy abecedy nevhodně

```
def prevod_morse(text):
    vystup = ''
    for i in range(len(text)):
        if text[i] == 'A': vystup += '.-|'
        elif text[i] == 'B': vystup += '-...|'
        elif text[i] == 'C': vystup += '-.-.|'
        elif text[i] == 'D': vystup += '-..|'
        # atd
    return vystup
```

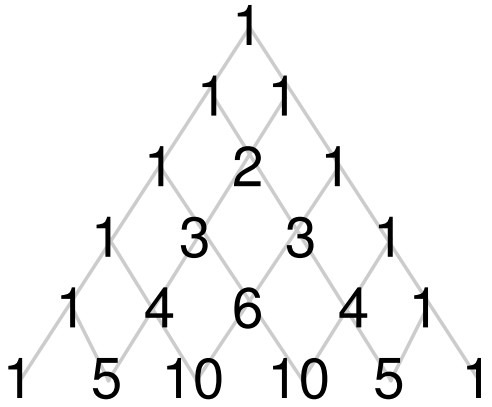
Převod do Morseovy abecedy: využití seznamu

```
morse = ['.-', '-...',  '-.-.', '-..'  ] # atd

def prevod_morse(text):
    vystup = ''
    for i in range(len(text)):
        if ord('A') <= ord(text[i]) <= ord('Z'):
            c = ord(text[i]) - ord('A')
            vystup += morse[c] + '|'
    return vystup
```

(ještě lepší řešení: využití slovníku)

Pascalův trojúhelník



$$\binom{0}{0}$$
$$\binom{1}{0}\ \binom{1}{1}$$
$$\binom{2}{0}\ \binom{2}{1}\ \binom{2}{2}$$
$$\binom{3}{0}\ \binom{3}{1}\ \binom{3}{2}\ \binom{3}{3}$$

Explicitní vzorec

$$\binom{n}{k} = \frac{n!}{(n-k)!k!}$$

Rekurzivní vztah

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$$

Pascalův trojúhelník

```
def pascaluv_trojuhelnik(n):
    aktualni_radek = [ 1 ]
    for i in range(n):
        for x in aktualni_radek:
            print x,
        print
        dalsi_radek = [ 1 ]
        for i in range(len(aktualni_radek)-1):
            dalsi_radek.append(aktualni_radek[i] +\
                               aktualni_radek[i+1])
        dalsi_radek.append(1)
        aktualni_radek = dalsi_radek
```

Prvočísla

- dělitelné jen 1 a sebou samým
- předmět zájmu matematiků od pradávna, cca od 70. let i důležité aplikace (moderní kryptologie)
- problémy s prvočísly:
  - výpis (počet) prvočísel v intervalu
  - test prvočíselnosti
  - rozklad na prvočísla (hledání dělitelů)

Výpis prvočísel přímočaře

```
def vypis_prvocisel(kolik):
    n = 1
    while kolik > 0:
        if len(seznam_delitelu(n)) == 2:
            print n,
            kolik -= 1
        n += 1
```

Odbočka: test prvočíselnosti, kryptografie

Test prvočíselnosti:

- zkoušíme všechny možné dělitele od 2 do $n - 1$
- vylepšení:
  - dělíme pouze lichými čísly
  - dělíme pouze čísly tvaru $6k \pm 1$
  - dělíme pouze do $\sqrt{n}$

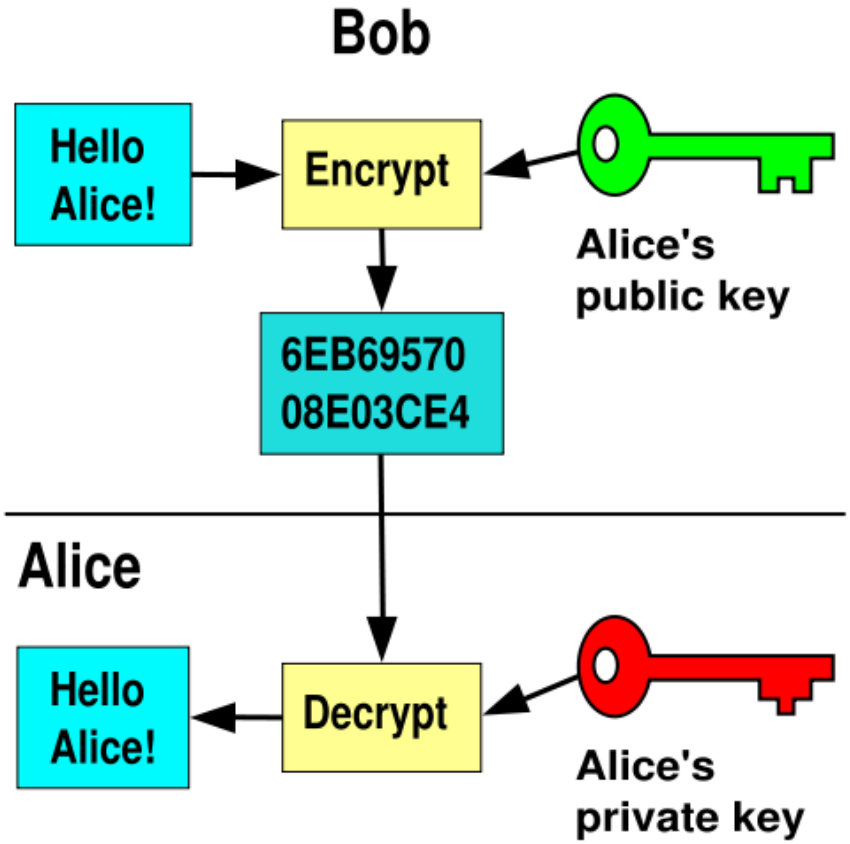
Test prvočíselnosti: chytřejší algoritmy

- náhodnostní algoritmy
- polynomiální deterministický algoritmus (objeven 2002)
- (vysoce) nad rámec tohoto kurzu
- umí se to dělat rychle

Rozklad na prvočísla

- rozklad na prvočísla = faktorizace
- naivní algoritmy:
  - průchod všech možných dělitelů
  - zlepšení podobně jako u testů prvočíselnosti
- chytřejší algoritmy:
  - složitá matematika
  - aktivní výzkumná oblast
  - neumí se to dělat rychle
  - max cca 200 ciferná čísla

Příklad aplikace: asymetrická kryptologie



http://en.wikipedia.org/wiki/Public-key_cryptography

Asymetrická kryptologie: realizace

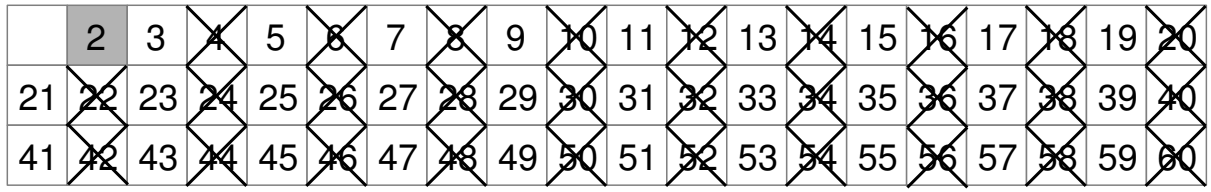
- jednosměrné funkce
  - jednoduché vypočítat jedním směrem
  - obtížné druhým (inverze)
  - ilustrace: míchání barev
- RSA (Rivest, Shamir, Adleman) algoritmus
  - jednosměrná funkce: násobení prvočísel (inverze = faktorizace)
  - veřejný klíč: součin velkých prvočísel
  - bezpečnost $\sim$ nikdo neumí provádět efektivně faktorizaci
  - využití modulární aritmetiky, Eulerovy věty, ...

Eratosthenovo síto

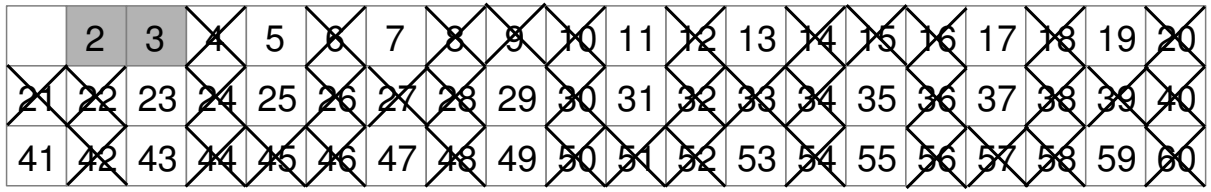
- problém: výpis prvočísel od 2 do n
- algoritmus: opakovaně provádíme
  - označ další neškrtnuté číslo na seznamu jako prvočíslo
  - všechny násobky tohoto čísla vyškrtni

Eratosthenovo síto

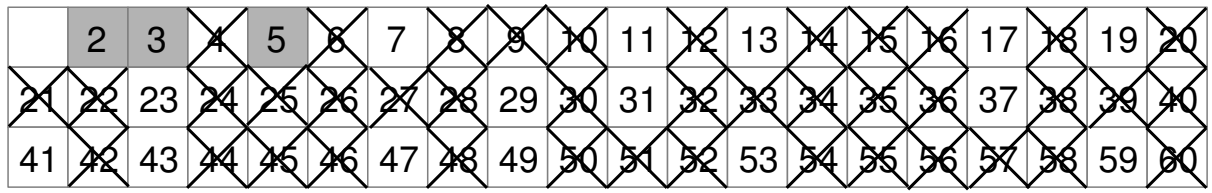
1. krok



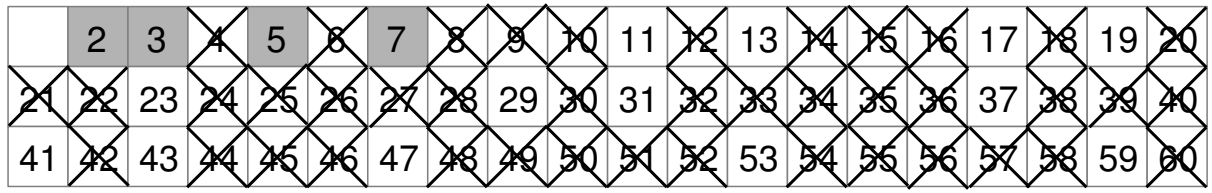
2. krok



3. krok



4. krok



Eratosthenovo síto

```
def eratosthenes(pocet):
    je_kandidat = [ 1 for i in range(pocet) ]
    for i in range(2, pocet):
        if je_kandidat[i]:
            print i,
            k = 0
            while k < pocet:
                je_kandidat[k] = 0
                k += i
```

Funkcionální prvky v Pythonu

- funkcionální programování
  - výpočet jako vyhodnocení matematické funkce
  - předmět IB015, jazyk Haskell
- Python obsahuje funkcionální prvky, např.
  - generátorová notace seznamů (list comprehension)
  - funkce map, reduce, filter
  - lambda výrazy

Funkcionální prvky v Pythonu – ukázka

```
n = 12
delitele = [ i for i in range(1, n+1) if n % i == 0 ]

print delitele
print map(str, delitele)
print map(lambda x: 'I'*x, delitele)
print filter(lambda x: x > 3, delitele)
print reduce(lambda x,y: x*y, delitele)
```

Příklady na rozmyšlení

- rozměňování mincí
  - klasické mince: 1, 2, 5, 10, 20, 50, 100, ...
  - zadané hodnoty mincí
- jednorozměrné piškvorky

Shrnutí

seznamy, řetězce:

- základní operace
- ukázky použití
- kryptografické příklady (historické) a souvislosti (moderní)
- příště: vyhledávání, řadicí algoritmy