

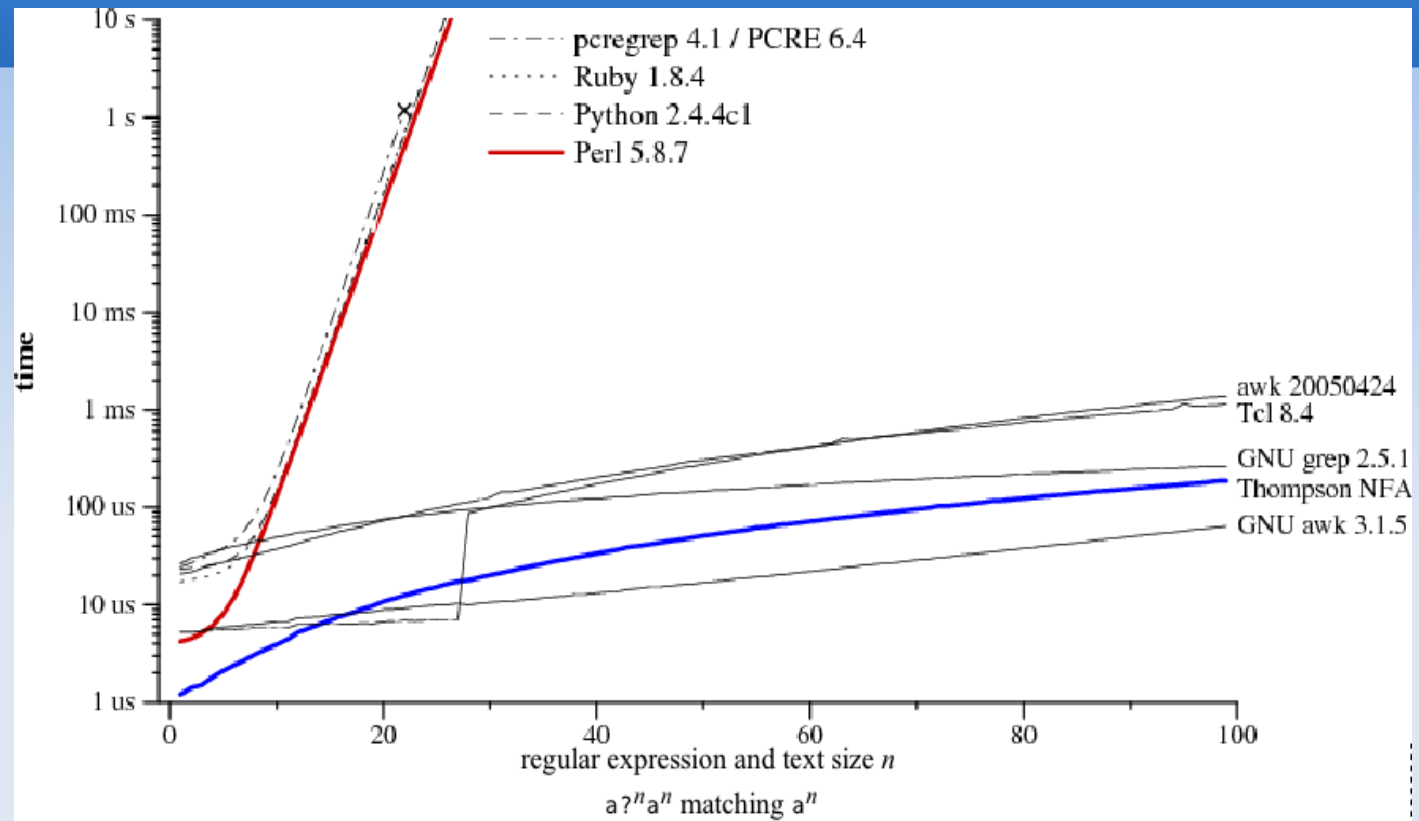
# *Moderné regulárne výrazy*

Tatiana Tóthová  
školiteľ RNDr. Michal Forišek, PhD.

# Cieľ

- *Spraviť prehľad nových konštrukcií používaných v moderných knižniciach s regulárnymi výrazmi (ako napr. look-ahead a look-behind assertions)*
- *Analyzovať tieto rozšírenia z hľadiska formálnych jazykov a prípadne tiež z hľadiska algoritmickej výpočtovej zložitosti*

# Motivácia



- *Formálne jazyky → prvé implementácie → Perl → nové konštrukcie pre lepšie a ľahšie používanie*
- *Sú to iba kozmetické úpravy?*

# Nové konštrukcie v regulárnych výrazoch (1)

- $*?$ ,  $+?$ ,  $??$  - funkcia rovnaká, berú čo najmenej

**Veta.**

*Regulárne jazyky sú uzavreté na operácie  $*?$ ,  $+?$ ,  $??$ .*

# Nové konštrukcie v regulárnych výrazoch (2)

- $(?=...)$  look-ahead,  $(?<=...)$  look-behind
- $(?!...)$  negatívny l-a,  $(?<!...)$  negatívny l-b

**Veta.** Regulárne jazyky sú uzavreté na l-a a l-b.

**Veta.** Bezkontextové jazyky nie sú uzavreté na l-a a l-b.

**Veta.** Kontextové jazyky sú uzavreté na l-a a l-b.

Pozn.: l-a môže obsahovať regex, l-b má fixnú dĺžku

# Nové konštrukcie v regulárnych výrazoch (3)

- `\m` spätná referencia
  - reprezentuje string, ktorý zachytil regex v m-tej dvojici zátvoriek
- **trieda E-regex**
  - *podmnožina kontextových jazykov*
  - *neporovnateľná s bezkontextovými jazykmi*
  - *uzavretá na homomorfizmus*
  - *nie je uzavretá na komplement, inverzný homomorfizmus, konečnú substitúciu, shuffle s regulárnym jazykom a prienik*

# Nové konštrukcie v regulárnych výrazoch (4)

- **trieda LE-regex**
  - *podmnožina kontextových jazykov*
  - *neporovnateľná s bezkontextovými jazykmi*
  - *uzavretá na nevymazávajúci homomorfizmus, prienik a komplement*
  - *nie je uzavretá na inverzný homomorfizmus*
  
- *(?) vymazávajúci homomorfizmus, konečná substitúcia, shuffle s regulárnym jazykom*

# Nové konštrukcie v regulárnych výrazoch (5) – čo ďalej?

- **trieda LE-regex**
  - *(?) nejaký model + pumpovacia lema*
  - *(?) vymazávajúci homomorfizmus, konečná substitúcia, shuffle s regulárnym jazykom*
- spätné referencie odkazujúce do look-ahead  
*( !!! look-ahead nerobí backtrack )*
- negatívne look-ahead, look-behind
- look-behind fixnej dĺžky



*Ďakujem za pozornosť*

*<https://github.com/Tatianka/bak>*