

## Regulární výrazy a lexikální analýza 1

(195.113.207.171:32)

Cíl: ukázat tradiční přístupy k analýze obsahu vstupního souboru.

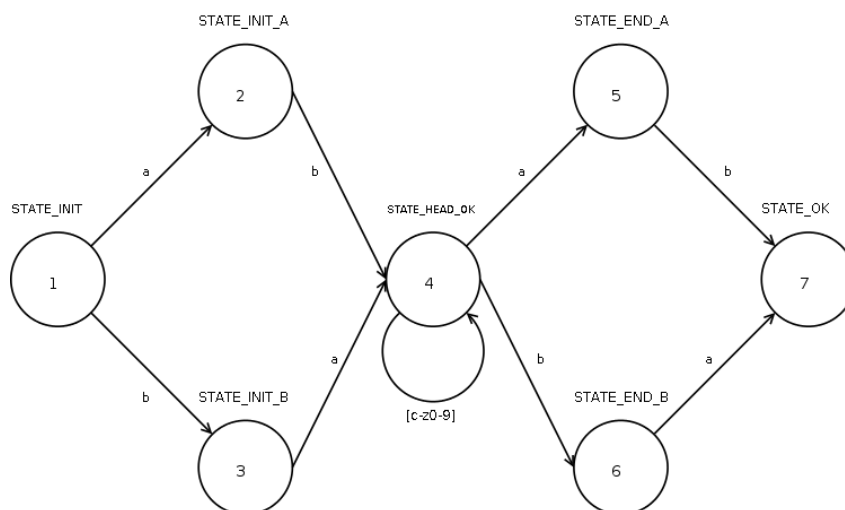
### Příklad 1: Filtry *grep*, *egrep*

Pomocí standardních příkazů operačního systému zpracujte text ze souboru *test1\_1.txt* tak, abyste z něj získali pouze následující lexikální elementy:

1. Slova, začínající písmenem, následovaným jedinou číslicí
2. Slova, obsahující písmena, případně čísla (*A,B,...,Z,a,b,...,z,0,...,8,9*)
3. Slova, obsahující sedm nebo osm malých písmen
4. Slova, obsahující výhradně písmena, včetně prázdného řetězce
5. Slova, začínající a končící slabikou 'ab' nebo 'ba', mezi nimiž mohou být písmena *c až z*, případně číslice
6. Všechna přirozená čísla s nejméně jednou cifrou

### Příklad 2: Konečný automat jako konvenční program

Bez pomoci knihoven RE napište program v jazyce C, fungující jako konečný automat, který akceptuje pouze jazyk z bodu 5 předchozího příkladu, tj. *slova, začínající a končící slabikou 'ab' nebo 'ba', mezi nimiž mohou být písmena c až z, případně číslice*. Inspirovat se přitom můžete následujícím diagramem:



Svá řešení můžete založit na nedokončeném programu ze souboru *main\_draft.c*.

Příkazy z příkladu 1 uložte do textového souboru. Ten společně s výsledným C-kódem příkladu 2 uložte do Vámi vytvořeného podadresáře:

`S:\ucitel\Voracek_Jan\XPJP\cviceni\2_3\<prijemni_nebo_login>`