

Wyrażenia regularne

Wyrażenia regularne (ang. *regular expressions*) to specjalny język służący do opisywania zbiorów ciągów. Z naszego punktu widzenia można go scharakteryzować jako zestaw symboli używanych do wyszukiwania ciągów znaków, tj. pojedynczych wyrazów, ich części lub związków kilkuwyrazowych.

Na co dzień wyrażen regularnych używamy (czasami nawet nieświadomie) wyszukując słów w edytorach tekstu, wyszukiwarkach internetowych czy słownikach. Język ten jest jednak najbardziej przydatny w przeszukiwaniu korpusów tekstów.

Oto lista najważniejszych operatorów wykorzystywanych w budowaniu wyrażen regularnych (lista ta jest oczywiście niekompletna i przedstawia tylko kilka najistotniejszych znaków):

- — dowolny znak;
- * — dowolna ilość wystąpień poprzedzającego operator znaku, w tym brak wystąpień;
- + — dowolna ilość wystąpień poprzedzającego operator znaku, ale co najmniej jedno;
- ? — jedno lub zero wystąpień poprzedzającego operator znaku;
- ^ — początek linii;
- \$ — koniec linii;
- | — alternatywa.

Oto kilka przykładów utworzonych za pomocą tych operatorów wyrażen regularnych oraz opis wyników, jakie zostaną przez nie zwrócone:

- ^lexic.* — wszystkie ciągi rozpoczynające się od “lexic”;
- *lexic\$ — wszystkie ciągi kończące się na “lexic”;
- *lexic.* — wszystkie ciągi zawierające część “lexic”;
- ^.ox\$ — wszystkie ciągi trzyliterowe, rozpoczynające się dowolnym znakiem i kończące na “ox”;
- (hat|cat) — wszystkie ciągi zawierające część “hat” lub “cat”;

Aby lepiej zgłębić temat proszę zapoznać się z poniżymi stronami o wyrażeniach regularnych: w języku polskim http://www.linux.net.pl/~wkotwica/doc/grep/grep_6.html lub w języku angielskim http://gnosis.cx/publish/programming/regular_expressions.html

a następnie spróbować swoich sił w ćwiczeniach <http://regex.sketchengine.co.uk/>

Proszę również zapoznać się z instrukcjami dwóch wyszukiwarek dla korpusu NKJP:

1. Poliqarp wyszukiwarka <http://nkjp.pl/poliqarp/>
instrukcja <http://nkjp.pl/poliqarp/help/plse3.html#x4-50003>
2. PELCRA wyszukiwarka <http://www.nkjp.uni.lodz.pl/>
instrukcja http://www.nkjp.uni.lodz.pl/help.jsp#skladnia_w_przykladach

Beata Wójtowicz