

Regular expressions overview

1 Overview

- Pattern matching
- Simple strings — regular expressions matching only themselves
- Generalization is possible via using special symbols
- Used to enhance corpus search

2 Syntax

2.1 Symbol patterns

- [...] — set [aoiuye] → any English vowel (exactly one!)
- [...-...] — set in an interval [a-zA-Z0-9] → a letter or a digit (exactly one!)
- [^...] — ‘anti’ set [^aoiuye] → any other symbol (exactly one!)
- . — any symbol k.lt → kalt, kblt, kclt, ...
- ...|... — alternative kalt|kält → kalt or kält
- (...) — group k(a|ä)lt → kalt or kält

2.2 Quantifiers

- ? — 0 or 1 (optional) kalt(er)? → kalt or kalter
- * — 0 or more [aoiuye]* → zero or more English vowels
- + — 1 or more [aoiuye]+ → at least one English vowel
- {a,b} — a number between a and b [aoiuye]1,4 → 1, 2, 3, or 4 English vowels

3 Examples

See <http://regexr.com/> to test your regular expressions.

Example 1: Forms of the word kalt: kalt, kält, kalter, kältester, kältesten

- k(a|ä)lt(er)? → kalt, kält, kalter, kälter
- k(a|ä)lt(er|este(r|n))? → kalt, kält, kalter, kälter, kaltester, kaltesten, kältester, kältesten