

Syntax (V)

Regular Expressions (cont.)

- **Special characters** in regular expressions (cont.):

- **()** : **group tokens** (*different from EBNF*)
 - `th(e|is)` : the, this
- **+** : **match the prior expression one or more times**
 - html tags: `<html> </html>`, `<h1></h1>`, `<div id="block1"></div>`
 - `<[A-Za-z][A-Za-z0-9]*>`: matches HTML tags without any attributes
 - `<[A-Za-z0-9]+>`: matches HTML tags without any attributes, but can have invalid tag like `<1>`
 - `<[^<>]+>` : matches HTML tags without regard to attributes
- **{min, max}** : specify **how many times a token can be repeated**, min ≥ 0 minimum number of matches, max $\geq$ min maximum number of matches. If **{min, }** the maximum number of matches is infinite. **{min}** repeat exactly min times.
 - `{0, }` same as `*`, `{1, }` same as `+`
 - `^[1-9][0-9]{3}$`: matches a number between 1000 and 9999
 - `^[1-9][0-9]{2,4}$`: matches a number between 100 and 99999