

프로그래밍 언어의 기초

정규표현식

강 범 일
(kangbeomil@gmail.com)

정규 표현식(Regular Expression)

- 문자열의 패턴을 구성하는 식

기호	설명
^	행의 시작
\$	행의 끝
.	하나의 문자. 단, 개행(new line) 문자는 제외.
[...]	대괄호 안에 있는 문자 중 하나의 문자
[^...]	대괄호 안에 있는 문자들이 아닌 문자
*	앞에 있는 문자가 0회 이상 반복
+	앞에 있는 문자가 1회 이상 반복
?	앞에 있는 문자가 있거나 없거나
{n}	앞에 있는 문자가 n회 반복
{n,}	앞에 있는 문자가 n회 이상 반복
{ n, m}	앞에 있는 문자가 n회 이상, m회 이하 반복

정규 표현식(Regular Expression)

기호	설명
a b	a 또는 b
\ww	단어를 구성하는 문자
\WW	단어를 구성하지 않는 문자
\ws	공백 문자
\WS	공백 문자가 아닌 문자
\wd	숫자
\WD	숫자가 아닌 문자
\wb	단어 경계
\WB	단어 경계가 아닌 것
\wn	개행(new line) 문자
\wt	탭 문자
(...)	정규 표현식을 grouping하고, grouping한 정규 표현식으로 검색한 결과를 기억
\w1...\w9	grouping한 정규표현식으로 검색한 결과 지칭

정규 표현식(Regular Expression)

- 파일: regex_sample.txt
 - '정상'으로 시작하는 어절 검색하기
 - 예) 정상, 정상은, 정상들, 정상회담
 - '정상'으로 시작하되 '정상회담', '정상회의'로 시작하지 않는 어절 검색하기
 - 어절 중간에 '정상'이 포함된 어절 검색하기
 - '숫자+마침표+공백' 패턴 제거하기
 - '0월 0일', '0/0' 형식으로 변환
 - 이메일 패턴 검색하기
 - '한 줄에 한 어절'이 나타나도록 문서 구조 변경
 - '한 줄에 한 글자'가 나타나도록 문서 구조 변경
 - '한 줄에 한 문장'이 나타나도록 문서 구조 변경