

# 정규 표현식 빠른 참조

패턴, 수량자, 그룹, 전방탐색, 플래그

## 기본 패턴

<b>메타 문자</b>
<b>.</b> 임의의 문자 (개행 제외)
<b>^</b> 문자열 / 줄의 시작
<b>\$</b> 문자열 / 줄의 끝
<b>*</b> 앞의 것이 0회 이상
<b>+</b> 앞의 것이 1회 이상
<b>?</b> 앞의 것이 0회 또는 1회 (선택)
<b>\</b> 메타 문자 이스케이프

<b>리터럴 매칭</b>
hello # matches "hello" exactly
a.c # matches "abc", "alc", "a-c", etc.
\\.txt # matches literal ".txt"

## 문자 클래스

<b>대괄호 표현식</b>
<b>[abc]</b> a, b, 또는 c 매칭
<b>[^abc]</b> a, b, c를 제외한 모든 것 매칭
<b>[a-z]</b> 소문자
<b>[A-Z]</b> 대문자
<b>[0-9]</b> 숫자
<b>[a-zA-Z0-9]</b> 영숫자

<b>단축 클래스</b>
<b>\d</b> 숫자 [0-9]
<b>\D</b> 비숫자 [^0-9]
<b>\w</b> 단어 문자 [a-zA-Z0-9_]
<b>\W</b> 비단어 문자
<b>\s</b> 공백 [\t\n\rf]
<b>\S</b> 비공백

## 수량자

<b>탐욕적 수량자</b>
<b>*</b> 0회 이상 (탐욕적)
<b>+</b> 1회 이상 (탐욕적)
<b>?</b> 0회 또는 1회 (탐욕적)
<b>{n}</b> 정확히 n회
<b>{n,}</b> n회 이상
<b>{n,m}</b> n회에서 m회 사이

<b>게으른 수량자</b>
<b>*?</b> 0회 이상 (게으른 / 비탐욕적)
<b>+?</b> 1회 이상 (게으른)
<b>??</b> 0회 또는 1회 (게으른)
<b>{n,m}?</b> n회에서 m회 사이 (게으른)

<b>탐욕적 vs 게으른</b>
<.+> # greedy: "<b>bold</b>"
<.+?> # lazy: "<b>"

## 앵커

<b>^</b> 문자열 시작 (^m` 플래그로 줄 시작)
<b>\$</b> 문자열 끝 (^m` 플래그로 줄 끝)
<b>\b</b> 단어 경계
<b>\B</b> 비단어 경계
<b>\A</b> 문자열 시작 (^m` 영 앞임)
<b>\Z</b> 문자열 끝 (^m` 영 앞임)

<b>앵커 예시</b>
Hello # starts with "Hello"
worlds # ends with "world"
\bword\b # "word" as whole word
\bword\b # "word" inside another word

## 그룹 및 교대

<b>캡처 그룹</b>
(abc) # capture group: match "abc"
(a b c) # alternation: a or b or c
(cat dog) # match "cat" or "dog"
(\d{3})-(\d{4}) # groups: "123-4567"

<b>그룹 유형</b>
<b>(pattern)</b> 캡처 그룹
<b>(?:pattern)</b> 비캡처 그룹
<b>(?P&lt;name&gt;pat)</b> 이름 있는 그룹 (Python)
<b>(?&lt;name&gt;pat)</b> 이름 있는 그룹 (JS, .NET)
<b>\1 \2</b> 그룹 1, 2 역참조
<b>a b</b> 교대: a 또는 b

## 전방탐색 및 후방탐색

<b>(?=pattern)</b> 긍정적 전방탐색
<b>(?!pattern)</b> 부정적 전방탐색
<b>(?&lt;=pattern)</b> 긍정적 후방탐색
<b>(?&lt;!pattern)</b> 부정적 후방탐색

<b>탐색 예시</b>
\d+(?= USD) # digits followed by " USD"
\d+(?! USD) # digits NOT followed by " USD"
(?<=s)\d+ # digits preceded by "s"
(?!\s)\d+ # digits NOT preceded by "s"

탐색은 문자를 소비하지 않고 위치를 매칭

## 일반 패턴

<b>\d{1,3}</b> (\.\d{1,3}){3}	IPv4 주소 (기본)
<b>[lw.+]+@[\w-]+\.[lw.]+</b>	이메일 (기본)
<b>https?://[\w.-/\-?&amp;#]=1+</b>	URL (기본)
<b>\((?\d{3})?\)[-.\s]?(\d{3})?[-.\s]?(\d{4})</b>	미국 전화번호
<b>\d{4}-\d{2}-\d{2}</b>	날짜 (YY-MM-DD)
<b>#[0-9a-zA-F]{6}</b>	16진수 색상 코드

이는 간소화된 패턴; 실제 사용 시 더 엄격한 검증 필요

## 플래그

<b>g</b> 전역: 첫 번째가 아닌 모든 매치 찾기
<b>i</b> 대소문자 무시 매칭
<b>m</b> 여러 줄: ^, \$ 가 줄 경계 매칭
<b>s</b> Dotall: . 이 개행도 매칭
<b>x</b> 상세: 공백 무시, 주석 허용
<b>u</b> 유니코드: 전체 유니코드 지원

## 언어별 플래그 사용

/pattern/gi	# JavaScript
re.compile(r"pat", re.I   re.M)	# Python
grep -iE "pattern"	# grep (extended)