



## 필드 추출 & 정규화

# 필드 추출

> Field : 이벤트를 자동 혹은 정규식으로 Parsing 한 Values의 Key

> Field를 추출을 해야 하는 이유

- 필드 추출은 데이터를 해석 하기 위한 첫 번째 단계
- 필드 추출은 인덱싱 과정에서 자동적으로 추출되거나 사용자가 임의로 추출 할 수도 있음
- 난해하거나 형태가 불규칙한 원시 데이터에 의미를 부여하여 데이터 해석을 도움

# 정규식과 정규 표현식

## > 정규식

- 정규식은 특정한 문자열을 쉽고 간단한 방법으로 찾아내기 위한 표현식으로, 특정 패턴에 일치하는 문자열의 집합을 필드화 시키기 위해 사용
- Splunk Enterprise 정규식은 PCRE(Perl 호환 정규식), 구체적으로는 PCRE C 라이브러리 사용
- 특정한 규칙을 가진 문자열의 집합을 표현하는 데 사용하는 형식 언어  
(텍스트 편집기와 프로그래밍 언어에서 문자열의 검색과 치환 지원)

## > 정규 표현식

- 주로 텍스트 탐색과 문자열 조작
- 하나의 문자와 일치
- 문자열의 일부분(substring)이나 전체 문자열의 치환을 지원

# 정규식

## > 정규식 용어 및 설명

용어	설명
literal	- 정규식을 사용하여 일치하는 항목을 검색할 정확한 문자 텍스트
regular expression	- 리터럴과 대조하여 검색하기 위해 사용되는 패턴을 정의하는 메타문자
groups	- 정규식을 사용하면 정규식 문자를 묶는 데 사용되는 괄호 유형으로 구별되는 여러 그룹을 분류할 수 있음 - 일반적으로 괄호는 일치 또는 캡처 그룹, 원자 그룹 및 룩어라운드에 사용하고, 대괄호는 문자 클래스를 정의하는 데 사용 - 중괄호는 반복을 정의하는 데 사용하고, 꺾쇠 괄호는 명명된 캡처 그룹을 정의하는 데 사용하며, 양쪽 대괄호는 Splunk Enterprise 전용 모듈형 regex 식에 사용
character class	- 대괄호로 묶은 정규식 문자로, 일치하는 문자열을 찾는 데 사용 - 대문자에 일치시키려면 [A-Z]처럼 하이픈을 사용하여 범위를 정의
character type	- 문자 유형은 와일드카드와 마찬가지로 특정 리터럴 일치를 간략하게 나타냄 - 예를 들어, 마침표(.)는 모든 문자와 일치하고, ww는 밑줄을 포함하는 단어 또는 영숫자와 일치함
anchor	- 앵커는 특히 return wr 및 newline wn 등과 같은 텍스트 형식 지정 위치와 일치하는 문자 유형
alternation	- 교체란 정규식에 대체 일치 패턴을 입력하는 것 - 파이프 문자( )를 사용하여 전체 정규식을 포함할 수 있는 대체 패턴을 구분함 예) grey gray는 "grey" 또는 "gray"와 일치합니다.
quantifiers, or repetitions	- 수량자( *, +, ? )는 그룹이 리터럴 패턴과 일치되는 방법을 정의하는 데 사용됨 예) *는 0 이상과 일치하고 +는 1 이상과 일치하고 ?는 0 또는 1과 일치함
back references	- 역참조는 나중에 사용하기 위해 다시 호출할 수 있는 리터럴 그룹 - 역참조는 달러 기호(\$)와 (0이 아닌) 숫자로 나타냄

# 정규식

## > 정규식 용어 및 설명

### 1. 문자 유형

용어	설명
.	새 라인이 아닌 모든 문자
\w	_을 포함한 알파벳 문자
\W	알파벳이 아닌 문자
\s	공백 문자(\t,\n,\r,\f)
\S	공백 문자가 아닌 것
\d	숫자 하나
\D	숫자가 아닌 문자

### 2. 그룹 및 연산자

용어	설명
()	그룹
[]	문자 클래스 정의
{n}	표현식을 n번 매치
<>	그룹 정의
*	0번 또는 그 이상의 매치 (Closure 연산)
+	1번 또는 그 이상의 매치
?	0또는 1번의 매치
	문자 앞 부분 혹은 뒷 부분 선택 (Union 연산)
^	패턴의 시작을 표시
\$	패턴의 끝을 표시





필드추출(Web)

# 필드추출(Web)

> 필드 추출에 사용할 필드 명

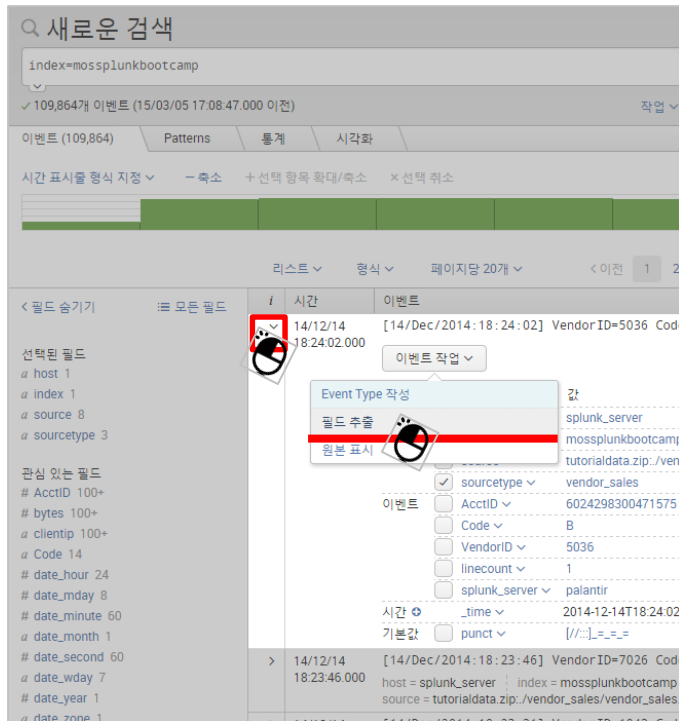
USER\_IP , DATE , METHOD , URL ,

ARG1 , STATUS , BYTES , RE\_URL , Browser

# 필드추출(Web)

## > 필드 추출의 예

### 1. ✓ 버튼 > 이벤트작업 > 필드추출



새로운 검색

index=mossplunkbootcamp

✓ 109,864개 이벤트 (15/03/05 17:08:47.000 이전)

이벤트 (109,864) Patterns 통계 시각화

시간 표시를 형식 지정 - 축소 + 선택 항목 확대/축소 × 선택 취소

리스트 형식 페이지당 20개 < 이전 1 2

< 필드 숨기기 모든 필드

선택된 필드

- a host 1
- a index 1
- a source 8
- a sourcetype 3

관심 있는 필드

- # AcctID 100+
- # bytes 100+
- a clientip 100+
- a Code 14
- # date\_hour 24
- # date\_mday 8
- # date\_minute 60
- a date\_month 1
- # date\_second 60
- a date\_wday 7
- # date\_year 1
- a date\_zone 1

이벤트

14/12/14 [14/Dec/2014:18:24:02] VendorID=5036 Code=1

이벤트 작업

Event Type 작성

필드 추출

원본 표시

source=splunk\_server index=mossplunkbootcamp

source=tutorialdata.zip vendor\_sales

이벤트

AcctID 6024298300471575

Code B

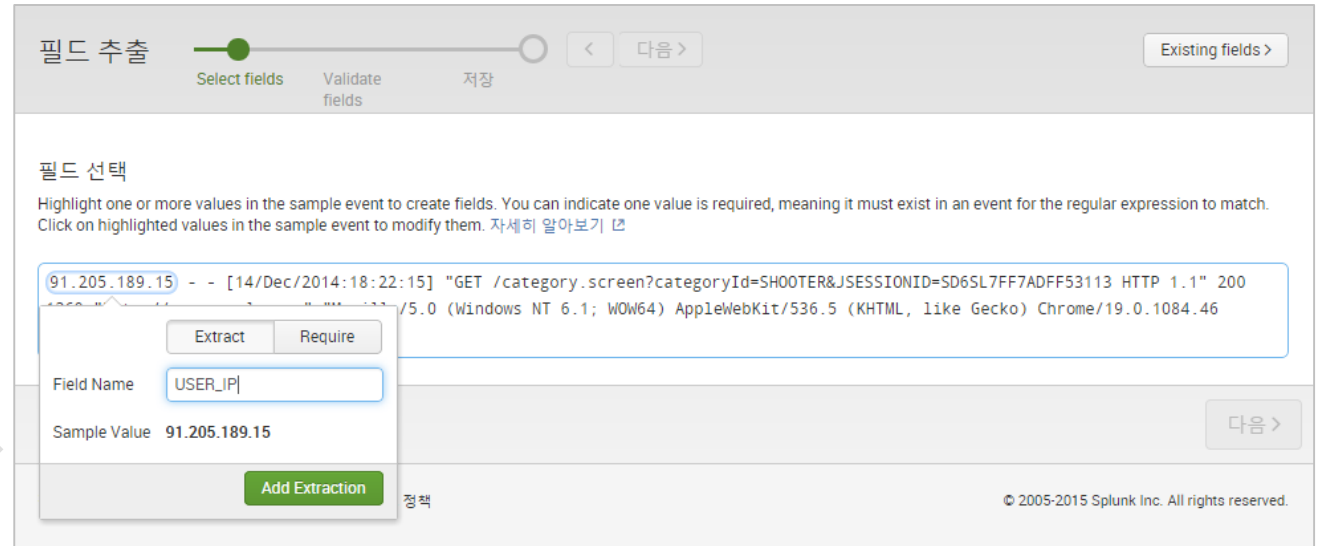
VendorID 5036

splunk\_server palantir

시간 2014-12-14T18:24:02

기본값 punct [/./-:]\_=-=

### 2. 추출 범위를 Drag 하고 필드명 입력



필드 추출

Select fields Validate fields 저장 < 다음 > Existing fields >

필드 선택

Highlight one or more values in the sample event to create fields. You can indicate one value is required, meaning it must exist in an event for the regular expression to match. Click on highlighted values in the sample event to modify them. 자세히 알아보기 >

91.205.189.15 - - [14/Dec/2014:18:22:15] "GET /category.screen?categoryId=SHOOTER&JSESSIONID=SD6SL7FF7ADFF53113 HTTP 1.1" 200 1369 "http://www.google.com" "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5" 779

Extract Require

Field Name USER\_IP

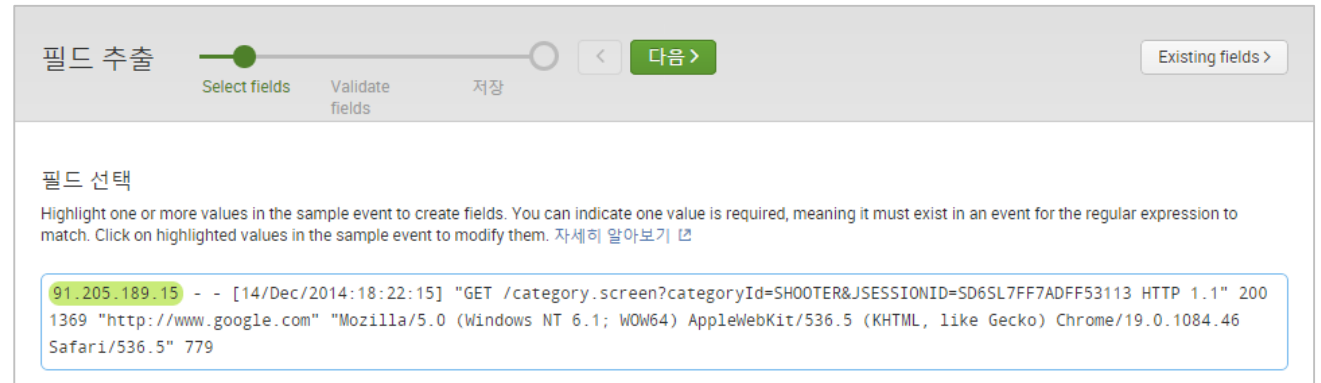
Sample Value 91.205.189.15

다음 >

Add Extraction 정책

© 2005-2015 Splunk Inc. All rights reserved.

### 3. 추출된 필드 확인



필드 추출

Select fields Validate fields 저장 < 다음 > Existing fields >

필드 선택

Highlight one or more values in the sample event to create fields. You can indicate one value is required, meaning it must exist in an event for the regular expression to match. Click on highlighted values in the sample event to modify them. 자세히 알아보기 >

91.205.189.15 - - [14/Dec/2014:18:22:15] "GET /category.screen?categoryId=SHOOTER&JSESSIONID=SD6SL7FF7ADFF53113 HTTP 1.1" 200 1369 "http://www.google.com" "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5" 779



# 필드추출(Web)

## > 필드 추출의 예

### 4. 추출 계속

필드 추출

Select fields

Validate fields

저장

<

다음 >

Existing fields >

필드 선택

Highlight one or more values in the sample event to create fields. You can indicate one value is required, meaning it must exist in an event for the regular expression to match. Click on highlighted values in the sample event to modify them. 자세히 알아보기

91.205.189.15 - - [14/Dec/2014:18:22:15] "GET /category.screen?categoryId=SHOOTER&JSESSIONID=SD6SL7FF7ADFF53113 HTTP 1.1" 200 1369 "http://www.myflowershop.com/category.screen?categoryId=SD6SL7FF7ADFF53113" Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5

Show Regular Expression

Field Name

Sample Value 14/Dec/2014:18:22:15

Extract

Require

Add Extraction

미리보기

If you see incorrect values in the next step.

to the set of sample events. Highlight its values to improve the extraction. You can remove incorrect values in

# 필드추출(Web)

## > 필드 추출의 예

### 5. 추출된 9개의 필드 확인

필드 추출

Select fields Validate fields 저장 < 다음 > Existing fields >

필드 선택

Highlight one or more values in the sample event to create fields. You can indicate one value is required, meaning it must exist in an event for the regular expression to match. Click on highlighted values in the sample event to modify them. 자세히 알아보기

91.205.189.15 - - [14/Dec/2014:18:22:15] "GET /category.screen?categoryId=SHOOTER&SESSIONID=SD6SL7FF7ADFF53113 HTTP 1.1" 200 1369 "http://www.google.com" "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5" 779

Show Regular Expression >

미리보기

If you see incorrect results below, click an additional event to add it to the set of sample events. Highlight its values to improve the extraction. You can remove incorrect values in the next step.

이벤트 USER\_IP DATE METHOD URL ARG1 STATUS BYTES RE\_URL Browser

✓ 1,000개 이벤트 (15/03/05 17:22:00.000 이전) 페이지당 20개 < 이전 1 2 3 4 5 6 7 8 9 ... 다음 >

필터 적용 샘플: 처음 1,000개의 이벤트 > All events > All Events Matches Non-Matches

	_raw	USER_IP	DATE	METHOD	URL
✓	91.205.189.15 - - [14/Dec/2014:18:22:16] "GET /oldlink?itemId=EST-14&SESSIONID=SD6SL7FF7ADFF53113 HTTP 1.1" 200 1665 "http://www.buttercupgames.com/oldlink?itemId=EST-14" "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5" 159	91.205.189.15	14/Dec/2014:18:22:16	GET	oldlink?itemId=EST-14&SESSIONID=SD6SL7FF7ADFF53113
✓	91.205.189.15 - - [14/Dec/2014:18:22:15] "GET /category.screen?categoryId=SHOOTER&SESSIONID=SD6SL7FF7ADFF53113 HTTP 1.1" 200 1369 "http://www.google.com" "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5" 779	91.205.189.15	14/Dec/2014:18:22:15	GET	category.screen?categoryId=SHOOTER&SESSIONID=SD6SL7FF7ADFF53113
✓	182.236.164.11 - - [14/Dec/2014:18:20:56] "GET /cart.do?action=addtocart&itemId=EST-15&productId=B5-A6-G09&SESSIONID=SD6SL8FF10ADFF53101 HTTP 1.1" 200 2252 "http://www.buttercupgames.com/oldlink?itemId=EST-15" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_7_4) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5" 506	182.236.164.11	14/Dec/2014:18:20:56	GET	cart.do?action=addtocart&itemId=EST-15&productId=B5-A6-G09&SESSIONID=SD6SL8FF10ADFF53101

### 6. 필드의 Value 확인 (1)

이벤트

USER\_IP

DATE

METHOD

URL

ARG1

STATUS

BYTES

RE\_URL

Browser

✓ 1,000개 이벤트 (15/03/05 17:22:00.000 이전)

페이지당 20개 ▾

필터

적용

샘플: 처음 1,000개의 이벤트 ▾

All events ▾

값 ▾	개수 ▾	%
200	862	86.633
503	27	2.714
400	20	2.010
408	19	1.910
406	18	1.809
404	17	1.709
500	14	1.407
505	14	1.407
403	4	0.402

### 7. 필드의 Value 확인 (2)

이벤트

USER\_IP

DATE

METHOD

URL

ARG1

STATUS

BYTES

RE\_URL

Browser

✓ 1,000개 이벤트 (15/03/05 17:22:00.000 이전)

페이지당 20개 >

필터

적용

샘플: 처음 1,000개의 이벤트 >


All events >

값	개수	%
Mozilla/5.0	706	70.955
Mozilla/4.0	247	24.824
Opera/9.20	26	2.613
Opera/9.01	13	1.307
Googlebot/2.1	3	0.302

# 필드추출(Web)

## > 필드 추출의 예

### 8. 필드 추출 최종 확인

필드 추출  Existing fields >

저장

Name the extraction and set permissions.

Extractions Name

소유자

앱

권한

---

Sourcetype

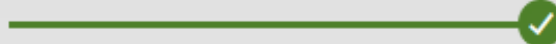
Sample event 

```
91.205.189.15 - - [14/Dec/2014:18:22:15] "GET /category.screen?categoryId=SHOOTER&JSESSIONID=SD6SL7FF7ADFF53113 HTTP/1.1" 200 1369 "http://www.google.com" "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5" 779
```

필드

정규식

### 9. 필드 추출 완료 화면

필드 추출 

Success!

You have extracted additional fields from your data (sourcetype=access\_combined\_wcookie).  
Edit your field extractions at any time by going to [Field Extractions](#).

What would you like to do next?

- [Explore the fields I just created in Search](#)
- [Extract more fields](#)

# 필드추출(Web)

## > 검색창에서 필드 확인

### 10. 모든 필드 > 필드 선택 > 선택된 필드에서 확인

새로운 검색

index=mossplunkbootcamp

전제 시간

109,864개 이벤트 (15/03/05 17:32:05.000 이전)

작업

스마트 모드

이벤트 (109,864) Patterns 통계 시각화

시간 표시줄 형식 지정 축소 선택 항목 확대/축소 선택 취소

월당 1월

리스트 형식 페이지당 20개

< 이전 1 2 3 4 5 6 7 8 9 ... 다음 >

< 필드 숨기기 모든 필드

선택된 필드

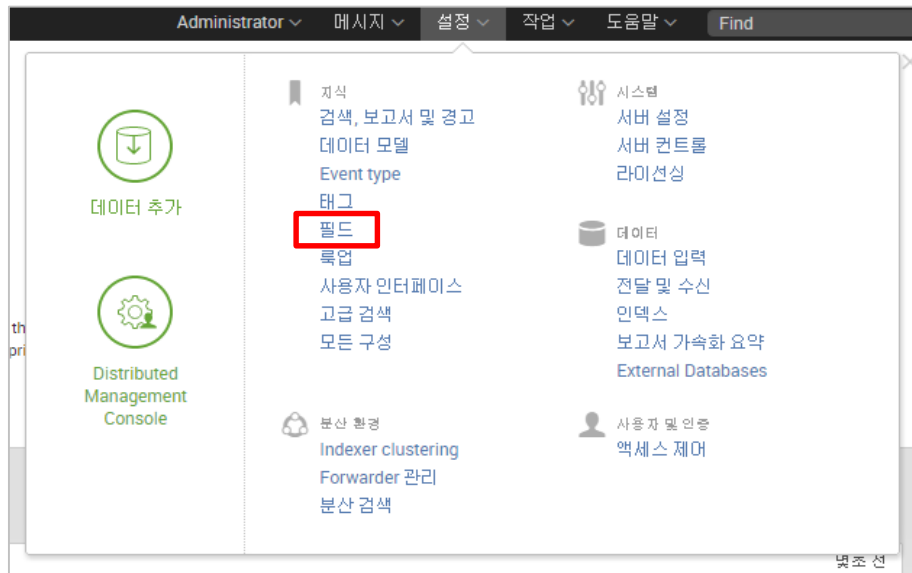
- ARG1 1
- BYTES 100+
- DATE 100+
- host 1
- index 1
- METHOD 2
- RE\_URL 100+
- source 8
- sourcetype 3
- STATUS 9
- URL 100+
- USER\_IP 100+

i	시간	이벤트
>	14/12/14 18:24:02.000	[14/Dec/2014:18:24:02] VendorID=5036 Code=B AcctID=6024298300471575 host = splunk_server index = mossplunkbootcamp source = tutorialdata.zip:/vendor_sales/vendor_sales.log sourcetype = vendor_sales
>	14/12/14 18:23:46.000	[14/Dec/2014:18:23:46] VendorID=7026 Code=C AcctID=8702194102896748 host = splunk_server index = mossplunkbootcamp source = tutorialdata.zip:/vendor_sales/vendor_sales.log sourcetype = vendor_sales
>	14/12/14 18:23:31.000	[14/Dec/2014:18:23:31] VendorID=1043 Code=B AcctID=2063718909897951 host = splunk_server index = mossplunkbootcamp source = tutorialdata.zip:/vendor_sales/vendor_sales.log sourcetype = vendor_sales
>	14/12/14 18:22:59.000	[14/Dec/2014:18:22:59] VendorID=1243 Code=F AcctID=8768831614147676 host = splunk_server index = mossplunkbootcamp source = tutorialdata.zip:/vendor_sales/vendor_sales.log sourcetype = vendor_sales
>	14/12/14 18:22:48.000	[14/Dec/2014:18:22:48] VendorID=1239 Code=K AcctID=5822351159954740 host = splunk_server index = mossplunkbootcamp source = tutorialdata.zip:/vendor_sales/vendor_sales.log sourcetype = vendor_sales

# 필드 관리

## > 필드 관리

### 1. 설정 버튼을 눌러 필드를 클릭



### 2. 필드 관리 화면에서 필드 추출 클릭

필드	
필드 추출에 대한 권한을 보거나 편집 및 설정합니다. 이벤트 워크플로 작업 및 필드 별칭을 정의합니다. sourcetype 이름을 변경합니다.	
유형	작업
<b>필드 별칭</b> 필드 이름을 편집하거나 하나 이상의 별칭을 추가합니다.	새로 추가
<b>계산 필드</b> 하나 이상의 계산 필드를 편집하거나 추가합니다.	새로 추가
<b>필드 추출</b> 모든 필드 추출을 보거나 편집합니다. 새 필드 추출을 추가하고 권한을 업데이트합니다.	새로 추가
<b>필드 변환</b> 변환을 사용하는 필드 추출에 대한 변환을 편집하거나 추가합니다.	새로 추가
<b>Sourcetype 이름 변경</b> source type 이름을 변경합니다. 다중 source type이 동일한 이름을 공유할 수 있습니다.	새로 추가
<b>워크플로 작업</b> 워크플로 작업 편집 또는 추가	새로 추가

# 필드 관리

## > 필드 추출의 예

### 1. 설정 버튼을 눌러 필드를 클릭

필드 추출  
필드 > 필드 추출

앱 컨텍스트 Search & Reporting (search) 소유자 모두

☐ 이 앱 컨텍스트에서 만든 개체만 표시 [자세히 알아보기](#)

[새로 만들기](#) [Open Field Extractor](#)

1개 항목 중 1-1 표시

이름	유형	추출/변환	소유자	App	공유 중	상태	작업
access_combined_wcookie : EXTRACT-USER_IP,DATE,METHOD,URL,ARG1,STATUS,BYTES,RE_URL,Browser	Inline	^(?P<USER_IP>[^\s]*)(?P<DATE>[^\s]*)(?P<METHOD>\w+)(?P<URL>[^\s]*)(?P<ARG1>[^\s]*)(?P<STATUS>[^\s]*)(?P<BYTES>[^\s]*)(?P<RE_URL>[^\s]*)(?P<Browser>[^\s]*)	admin	search	비공개	권한	사용 가능   삭제

### 2. 필드에 대한 권한을 지정 할 수 있음

권한  
필드 > 필드 추출 > access\_combined\_wcookie : EXTRACT-USER\_IP,DATE,METHOD,URL,ARG1,STATUS,BYTES,RE\_URL,Browser > 권한

Object이(가) 나타나야 함  
☐ 비공개 유지 ☐ 이 앱만(search) ☒ 모든 앱

권한

역할	읽기	쓰기
모든 사용자	<input checked="" type="checkbox"/>	<input type="checkbox"/>
admin	<input type="checkbox"/>	<input type="checkbox"/>
can_delete	<input type="checkbox"/>	<input type="checkbox"/>
power	<input type="checkbox"/>	<input type="checkbox"/>
splunk-system-role	<input type="checkbox"/>	<input type="checkbox"/>
user	<input type="checkbox"/>	<input type="checkbox"/>

[취소](#) [저장](#)

### 3. 필드 추출을 수정 할 수 있음

access\_combined\_wcookie : EXTRACT-USER\_IP,DATE,METHOD,URL,ARG1,STATUS,BYTES,RE\_URL,Browser  
필드 > 필드 추출 > access\_combined\_wcookie : EXTRACT-USER\_IP,DATE,METHOD,URL,ARG1,STATUS,BYTES,RE\_URL,Browser

추출/변환 \*

^(?P<USER\_IP>[^\s]\*)(?P<DATE>[^\s]\*)(?P<METHOD>\w+)

필드 추출이 인라인인 경우 정규식을 입력하십시오. 필드 추출에서 변환(transform)을 사용하는 경우, 변환(transform) 이름을 지정하십시오.

[취소](#) [저장](#)

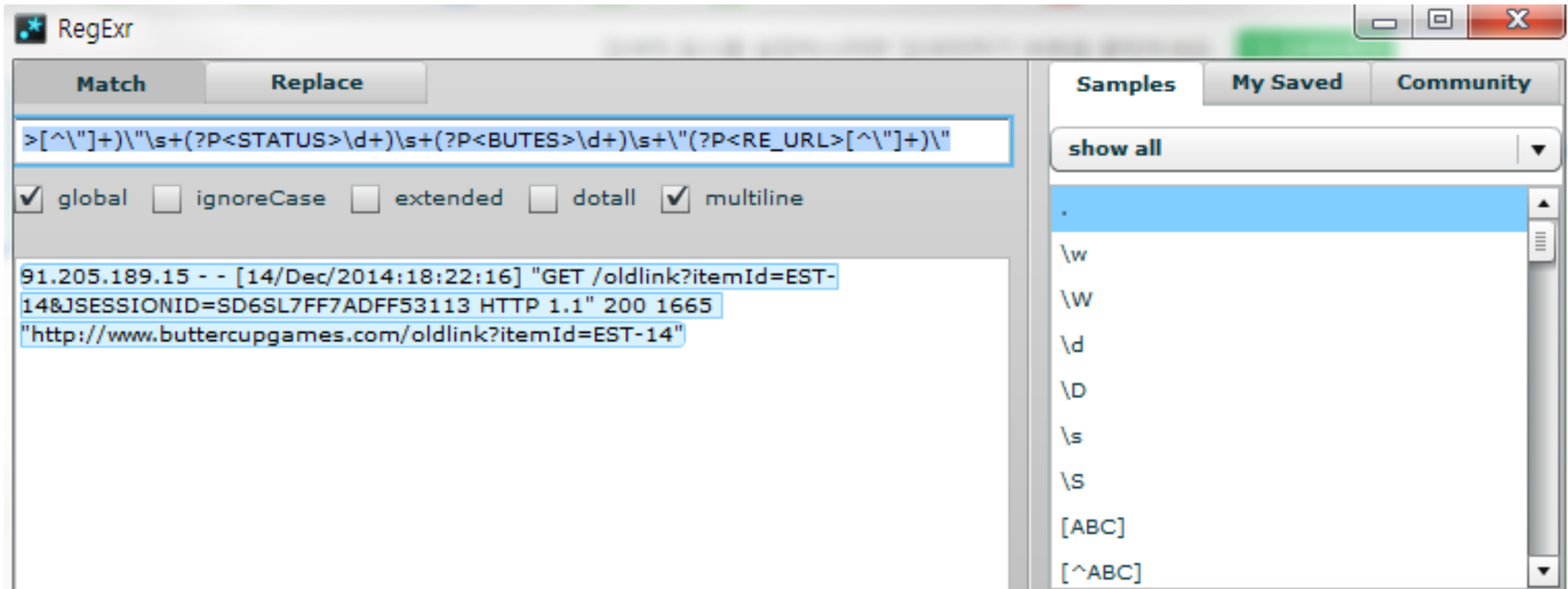




## 필드추출(REGX)

# 필드추출(REGX)

## > 필드 추출의 예



# 필드추출(REGX)

## > 로그 형태

- 정규식

```
^(?P<USER_IP>\S+)\s-\s-\s\[(?P<DATE>[^\]]+)\]\s+\\"(?P<METHOD>\w+)\s+(?P<URL>\S+)\s+(?P<ARG1>[^\"]+)\\"s+(?P<STATUS>\d+)\s+(?P<BYTES>\d+)\s+\\"(?P<RE_URL>[^\"]+)\\"s+\\"
```

- Log

```
89.11.192.18 - - [22/Mar/2013:11:46:18] "POST /oldlink?itemId=EST-26&JSESSIONID=SD6SL2FF5ADFF4953 HTTP
1.1" 200 3245 "http://www.myflowershop.com/oldlink?itemId=EST-26" "
```

# 필드추출(REGX)

## > 설명

1. 설정 버튼을 눌러 필드를 클릭 (IP 주소 추출 부분 설명)

```
^(?P<USER_IP>\d+\.\d+\.\d+\.\d+)
```

→ 시작(숫자.숫자.숫자.숫자)으로 구성된 범위를 USER\_IP로 명명)

2. CONF File 확인

경로 : \$Splunk/etc/apps/search/local/props.conf or \$Splunk/etc/system/local/props.conf

props.conf

```
[access_combined_wcookie]
```

```
EXTRACT-^(?P<USER_IP>\S+)\s-\s-\s-  
\\s\\[(?P<DATE>[^\]]+)\]\s+"(?P<METHOD>\w+)\s+(?P<URL>\S+)\s+(?P<ARG1>[^\"]+)\s+"(?P<STA  
TUS>\d+)\s+(?P<BYTES>\d+)\s+"(?P<RE_URL>[^\"]+)\s+"
```