

리스틀리 제품 설명서



리스틀리는 크롬, 엣지, 웨일 브라우저에서 설치할 수 있는 확장 프로그램입니다.
데이터 추출을 원하는 웹페이지에서 아이콘을 클릭하여 사용합니다.

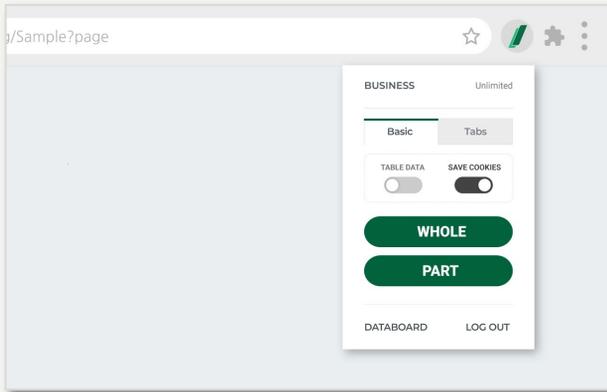
Listly Product Description Ver. Aug-04-2023
WWW.LISTLY.IO
Copyright 2023 LISTLY All rights reserved

LISTLY Product Description

INDEX

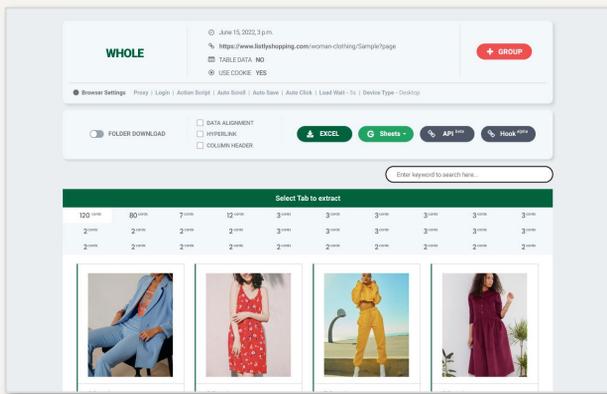


- 01 Listly 서비스 화면
- 02 Listly 용어 정리
- 03 확장 팝업 | Extension Popup
- 04 결과창 | Results Page
- 05 데이터 보드 | Data Board
- 06 설정 | Settings
- 07 대표적인 문제상황



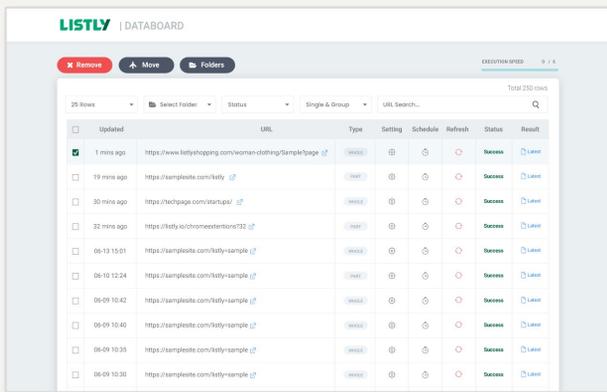
확장 팝업 | Extension Popup

데이터 추출 영역을 선택하는 팝업입니다.
리스트리 확장 아이콘을 클릭하면 등장합니다.



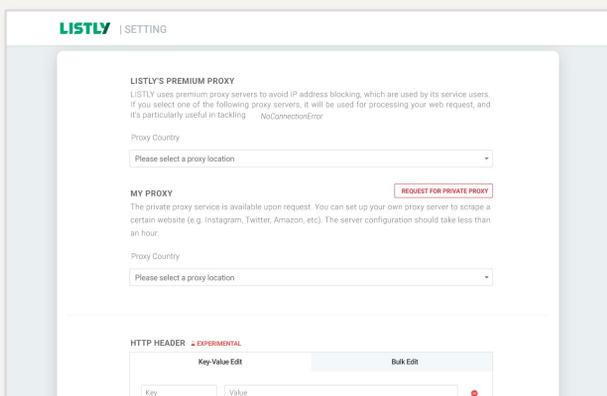
결과창 | Result Page

추출패널에서 영역을 선택한 후 이동하는 화면으로
웹페이지의 추출 결과를 확인할 수 있습니다. 데이터를
설정하고 엑셀파일로 내보냅니다.



데이터 보드 | Data Board

추출된 태스크들이 보관되는 공간입니다. 폴더를 생성하여
함께 추출할 태스크들을 묶을 수 있으며, 각 태스크의
설정화면으로 진입이 가능합니다.



설정 | Settings

프록시 / 접속기기 / 로그인 / 자동저장 / 자동스크롤 /
자동클릭 등 더욱 정확한 추출 결과를 위한 다양한 기능을
설정할 수 있습니다.

www.sampleurl.com/page1

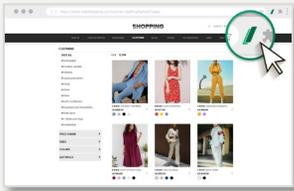
URL

웹페이지의 고유한 주소. URL을 통해 방문한 페이지를 스크랩 작업한 결과물이 하나의 태스크가 됩니다. 리스틀리에서는 추출 작업에 필요한 충전 자원을 의미하기도 합니다.

10-13 00:46	https://sampleite.com/listly	START	Success
10-11 17:45	Sample Site	WHOLE	100 SUCCESS FAIL DURING WAITING
06-13 15:01	https://sampleite.com/listly-sample	WHOLE	Success
06-10 12:24	https://sampleite.com/listly-sample	START	Success

태스크 (task)

데이터보드에 누적된 작업 단위를 말합니다. 싱글 태스크와 그룹 태스크가 있습니다.



타겟 페이지 (Target Page)

데이터를 추출할 대상 웹페이지를 말합니다.



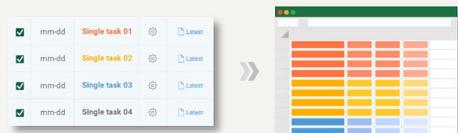
소스 페이지(Source Page), 그룹소스(Group Source)

그룹추출의 데이터 구조를 결정하는 추출 페이지



그룹추출 (Group Extraction)

다수의 URL을 입력해 여러 페이지를 하나의 엑셀시트로 추출하는 것을 의미합니다.



폴더 다운로드 (Folder Download)

다수의 태스크를 선택하여 하나의 엑셀시트로 내보내는 것을 의미합니다. 탭스 추출시 자동 적용됩니다.

1,000 cards

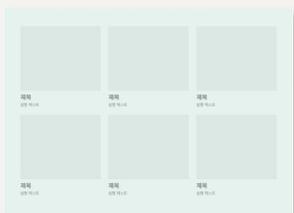
카드(Card)

리스틀리가 발견한 데이터 단위 영역으로, 숫자만큼의 유사 영역을 발견한 것입니다.

Select Tab to extract

탭(Tab)

같은 영역의 카드를 모아둔 화면입니다. 관련 데이터가 많이 포함되어 있는 순서로 제시됩니다.



목록 페이지(list page)

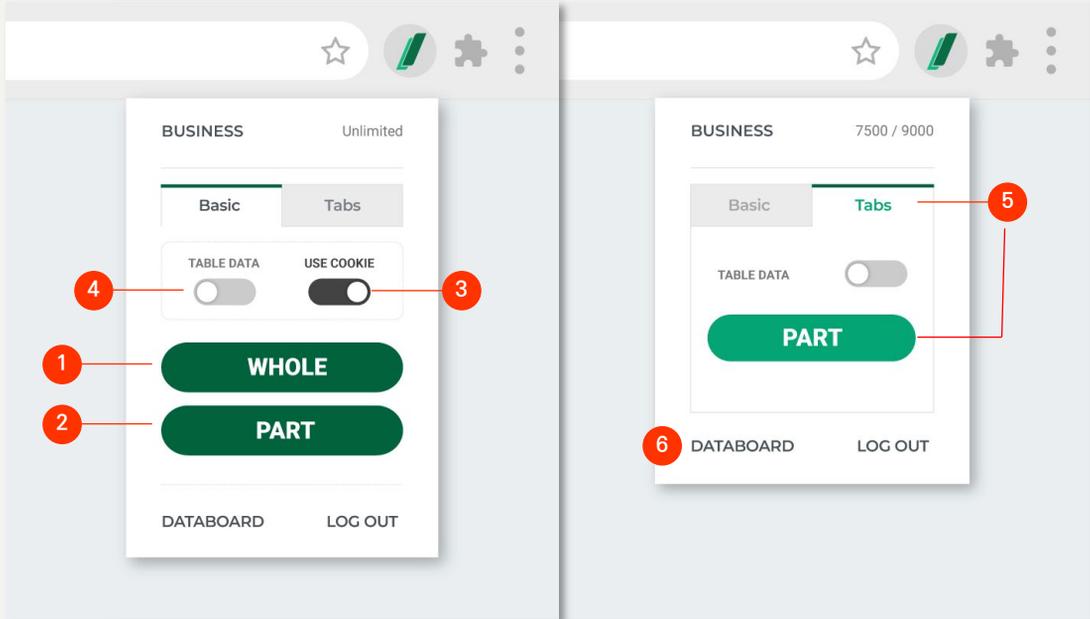
쇼핑몰의 썸네일 페이지와 같이 상세페이지로 진입하는 링크가 나열 되어있는 페이지입니다.



상세 페이지(detail page)

목록 페이지에서 진입할 수 있는 각각의 페이지입니다.

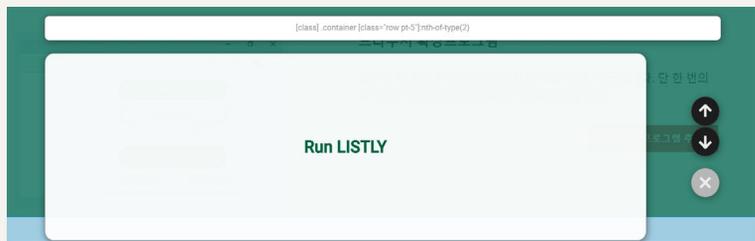
확장 팝업 | Extension Popup



1 WHOLE

화면 전체의 정보를 바탕으로 리스틀리가 자동으로 분석해 최적의 추출 영역을 추천해줍니다. 클릭 시 바로 결과창으로 이동합니다.

2 PART



LISTLY WHOLE보다 더 상세한 데이터 영역을 설정할 수 있습니다. 클릭 시 위와 같이 영역 선택 레이어가 등장합니다.

↑ : 선택 영역을 포함하는 더 큰 영역으로 확장합니다.

↓ : 현재 영역에서 포함하고 있는 바로 아래 영역을 선택합니다.

Run Listly : 선택 영역을 추출합니다.

3 USE COOKIE

웹사이트에 접속할 때 사용한 로그인 정보 등의 쿠키값을 저장합니다. 약 1시간 동안 유효합니다.

4 TABLE DATA

'표'처럼 생긴 데이터를 추출할 때 사용합니다. OFF가 기본값이지만 표처럼 생겼는데 결과가 만족스럽지 않을 경우 ON 상태로 추출을 다시 진행합니다.

5 Tabs-PART

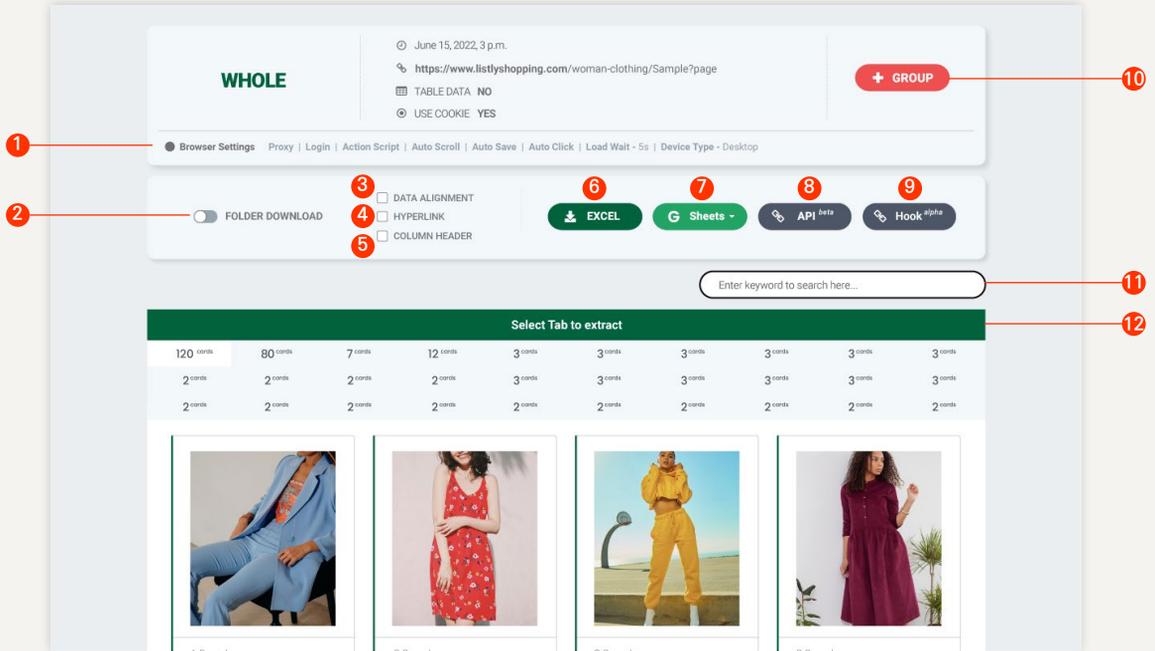
브라우저에 열려 있는 모든 탭에서 레이아웃이 일치하는 부분(PART)을 한 번에 추출 할 수 있는 기능입니다.

6 DATABOARD

선택 시 로그인 계정의 데이터보드로 바로 이동합니다.

*명칭을 클릭하면 관련 게시글로 이동합니다.

결과창 | Result Page



- 1

브라우저 세팅
Browser setting

추출한 페이지에 적용된 설정값을 확인할 수 있습니다. 클릭하여 설정화면으로 진입할 수 있습니다.
- 2

폴더 다운로드
Folder Download

토글버튼을 활성화 시키면 기존에 생성해둔 폴더 그룹이 나타납니다. 이 태스크가 특정 폴더에 포함되어 있을 경우 사용합니다. 탭(Tabs)로 추출했을 경우 각 페이지들의 데이터가 한 폴더에 담겨 자동으로 활성화됩니다.
- 3

데이터 정렬
Data Alignment

엑셀 내 데이터가 유사한 패턴끼리 자동으로 정렬되도록 하는 기능입니다. 해제하고 다운로드 할 경우 엑셀 파일에 빈 칸 없이 추출됩니다.
- 4

하이퍼링크
Hyperlink

각각의 데이터 값에 부여된 모든 하이퍼링크 값을 같이 추출합니다.
- 5

컬럼명
Column Header

체크 후 데이터를 내보내면 엑셀의 각 행 첫 번째 열에 행 이름이 생성됩니다.
- 6

엑셀 EXCEL

결과 데이터를 엑셀 파일로 다운로드합니다.
- 7

G Sheets

결과 데이터를 구글 스프레드시트로 내보내기 합니다.
- 8

API

API를 이용해 데이터를 json, csv 형태로 내보낼 수 있습니다.
- 9

HOOK

추출이 끝난 후의 결과를 사용자가 원하는 곳에 결과를 알려줍니다.
- 10

+그룹
+GROUP

현재 결과를 기반으로 연속 URL을 입력해 그룹추출을 할 수 있습니다.
- 11

검색바

키워드를 검색하여 추출할 데이터가 속해있는 탭 카드를 빠르게 찾을 수 있습니다.
- 12

탭 Tab

관련도가 높은 순으로 제시됩니다. 원하는 데이터가 있는 탭을 선택해 진행합니다.

*명칭을 클릭하면 관련 게시글로 이동합니다.

그룹 결과창 | Group Result Page

The screenshot shows the Group Result Page interface. At the top, there is a 'Sample site' link (1). Below it are buttons for 'GROUP EXCEL' (2), 'Sheets' (3), and 'GROUP API beta' (3), along with a 'LOGS' button. There are also checkboxes for 'ARRANGE CELL' and 'SHOW HYPERLINK'. A 'Group Source' bar (4) is below. Action buttons include 'REMOVE' (5), 'REFRESH' (6), and '+ URL' (7). The main table (8) has columns for 'Updated', 'URL', 'Status', 'Tab Matched', 'Screenshot', and 'Result'. A 'Screenshot' button (9) is also visible in the table header area.

Updated	URL	Status	Tab Matched	Screenshot	Result
09-27 17:37	https://samplesite.com/listly/page=01	Success	Y	-	Latest
09-27 17:42	https://samplesite.com/listly/page=02	Success	N	Screenshot	Latest
09-27 17:43	https://samplesite.com/listly/page=03	Success	N	Screenshot	Latest
09-27 17:40	https://samplesite.com/listly/page=04	Success	Y	-	Latest
09-27 17:39	https://samplesite.com/listly/page=05	Success	Y	-	Latest
09-27 17:41	https://samplesite.com/listly/page=06	Success	Y	-	Latest
09-27 17:43	https://samplesite.com/listly/page=07	Success	Y	-	Latest

- 그룹 이름**
Group Name
생성된 그룹의 이름을 수정할 수 있습니다.
- 그룹 엑셀**
Group Excel
생성된 그룹 태스크의 결과물을 엑셀로 다운받을 수 있는 버튼입니다.
- 로그 목록**
Logs
그룹추출 기록을 확인할 수 있습니다.
- 그룹 소스**
Group Source
해당 그룹의 소스로 활용되는 웹페이지의 정보가 미리보기 형태로 제시됩니다.
- 삭제**
Remove
그룹 추출 결과 및 데이터보드에서 선택된 태스크를 삭제합니다.
- 재추출**
Refresh
선택된 태스크의 데이터 추출을 다시 시도합니다.
- + URL**
현재 그룹에 추가하고 싶은 URL을 입력합니다.
- 탭매칭**
Tab Matching
데이터 구조의 일치 여부를 의미합니다.
- 스크린샷**
Screenshot
데이터 추출 실패 시 원인을 분석할 수 있는 화면입니다.

데이터 보드 | Data Board

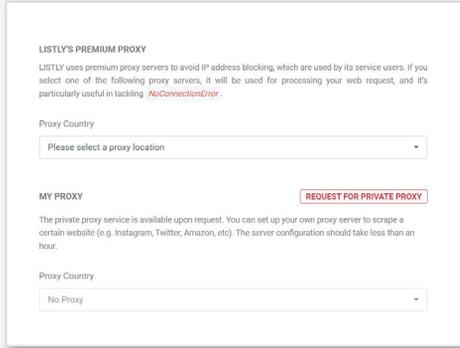
The screenshot shows the LISTLY DATABOARD interface. At the top, there are several action buttons: Remove (1), Move (2), Copy (3), Folders (4), Add, Remove, and Reset (5). A progress bar for 'EXECUTION SPEED' is shown on the right (6). Below the buttons, there are filters for '25 Rows', 'Select Folder', 'Status', 'Single & Group', 'Setting', and 'Schedule'. A search bar for 'URL, Group Search...' is present (7). The main table has columns: Updated, URL, Type, Setting, Schedule, Refresh, Status, and Result. The first row is selected, and its details are visible in the table below.

Updated	URL	Type	Setting	Schedule	Refresh	Status	Result
1 mins ago	https://www.listlyshopping.com/woman-clothing/Sample?page	WHOLE	⚙️	🕒	🔄	Success	Data
19 mins ago	https://samplesite.com/listly	PART	⚙️	🕒	🔄	Success	Data
30 mins ago	https://techpage.com/startups/	WHOLE	⚙️	🕒	🔄	Success	Data
32 mins ago	https://listly.io/chromeextensions732	PART	⚙️	🕒	🔄	Success	Data
06-13 15:01	https://samplesite.com/listly=sample	WHOLE	⚙️	🕒	🔄	Success	Data
06-10 12:24	https://samplesite.com/listly=sample	PART	⚙️	🕒	🔄	Success	Data
06-09 10:42	https://samplesite.com/listly=sample	WHOLE	⚙️	🕒	🔄	Success	Data
06-09 10:40	https://samplesite.com/listly=sample	WHOLE	⚙️	🕒	🔄	Success	Data
06-09 10:35	https://samplesite.com/listly=sample	WHOLE	⚙️	🕒	🔄	Success	Data

- 삭제** Remove 데이터보드에서 선택된 태스크를 삭제합니다.
- 이동** Move 선택한 태스크를 생성된 폴더로 이동합니다.
- 복사** Copy 선택한 태스크를 복제합니다. 복사만으로는 추출이 진행되지 않으며, 싱글 태스크의 일부 설정만 수정하거나, 동일한 설정에서 그룹 태스크에 속해있는 URL만 바꾸고 싶을 때 사용합니다.
- 폴더목록** Folders 폴더를 생성 / 수정 / 삭제할 수 있습니다.
- 초기화** Reset 추출속도 에러로 장시간 작업이 실행되지 않을 경우, 데이터보드의 모든 실행중, 대기중 작업들을 강제로 종료합니다.
- 실행속도**
Execution Speed 동시에 추출되는 웹페이지 개수를 설정해 추출속도를 조절합니다.
- 설정** Setting 각 태스크 혹은 그룹별로 추출 환경을 설정합니다.
- 스케줄** Schedule 설정한 시간대에 자동으로 새로운 데이터를 추출합니다.
- 재추출** Refresh 태스크를 다시 추출합니다. 특히 태스크에 설정을 추가했을 경우에는 재추출 버튼을 클릭해야 설정이 적용된 결과를 받을 수 있습니다.
- 상태** Status 해당 태스크의 추출 상태를 확인합니다. 표시 상태에는 완료 / 실패 / 추출중 / 대기중 이 있습니다.
- 결과** Result 해당 태스크의 결과창으로 이동합니다.

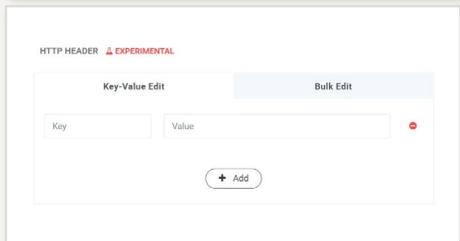
*명칭을 클릭하면 관련 게시글로 이동합니다.

설정 | Settings



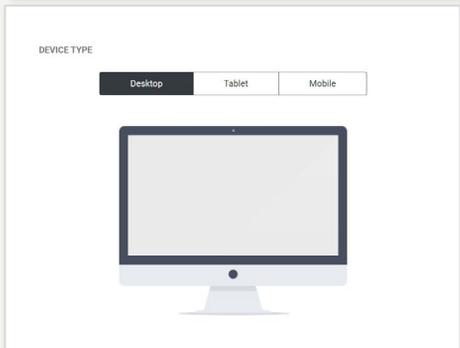
프록시 Proxy

웹사이트에서 차단당했거나, 국내 서버에서 접근이 어려운 웹페이지를 추출 할 때 접속 서버를 바꿔주는 기능입니다. 리스틀리가 보유한 자체 프록시 및 개별 구매한 프록시를 선택하여 적용할 수 있습니다.



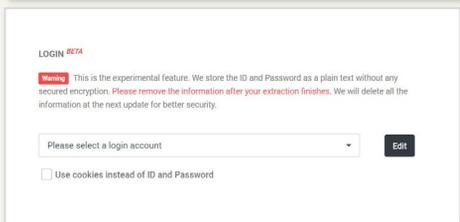
HTTP HEADER

봇의 접속을 제한해둔 웹사이트의 접속 성공률을 높이기 위한 추가 설정입니다.



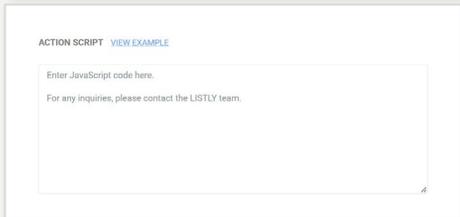
접속기기 종류 Device Type

모바일 뷰에서 보이는 페이지를 추출하고 싶은 경우에 사용합니다.



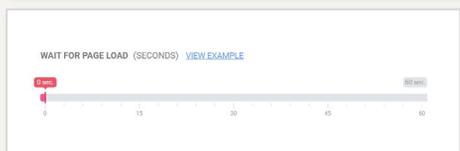
로그인 Login

해당 URL로 접속할 때 별도의 로그인 없이도 자동으로 로그인해 데이터를 추출할 수 있습니다.



액션스크립트 ActionScript

데이터 추출에 특정 동작이 필요한 경우 동작을 정의하는 JS 코드를 삽입할 수 있습니다. (개발자 전용)



잠시대기 Wait for page load

웹페이지에서 정보를 불러올 때 시간이 소요될 경우, 해당 페이지에 대기하는 시간을 설정합니다.

*명칭을 클릭하면 관련 게시글로 이동합니다.

설정 | Settings

AUTO SCROLL TO LOAD MORE DATA [VIEW EXAMPLE](#)

N Times to repeat Wait for

How many times scroll to bottom? Delay until next scrolling

Auto save while scrolling (beta) [VIEW EXAMPLE](#)

* Turning on this allows you to save your data automatically while scrolling.

자동스크롤 Auto Scroll

타킷 페이지에서 더 많은 상품 정보를 보기 위해 스크롤바를 계속 내려야 할 때 사용합니다.

AUTO CLICK TO LOAD MORE DATA [VIEW EXAMPLE](#)

Element to auto click (CSS Selector)

CSS Selector

N Times to repeat Wait for

How many times click? Delay until next clicking

자동클릭 Auto Click

타킷 페이지에서 더 많은 상품 정보를 보기 위해 특정 버튼을 계속 클릭해야 할 때 사용합니다.

스케줄 | Schedule

Current Time

Asia/Seoul
2022-12-22 17:16 (Thursday)

1 Timezone

Asia/Seoul (detected)

2 Period

How often scrape data?

3 Time

0:00

4 E-mail Alert

Cancel Save

- 1 **시간대** Timezone 데이터를 추출할 기준 시간대를 정합니다.
- 2 **반복 시점** Period 매일/매주/매월 단위로 해당 페이지 추출 반복 시점을 설정합니다.
- 3 **시간** Time 데이터 추출이 진행될 시간을 설정합니다.
- 4 **이메일 알람** E-mail Alert 이 기능을 활성화해둘 경우 가입시 인증한 메일로 추출 결과를 전송합니다.

대표적인 문제상황

데이터 추출 시 가장 많이 겪게 되는 문제 상황입니다.
더 많은 리스틀리 사용 방법은 [도움말 페이지](#)에서 확인하실 수 있습니다.



여러 페이지를 수집해야 하는 경우

그룹추출과 폴더그룹추출 중 어떤 방식으로 추출해야 할까요?
체크리스트를 따라 나의 상황에 맞는 추출 방식을 선택합니다.



로그인 설정이 필요한 경우

처음 추출 할 땐 성공했는데, 왜 재추출은 안될까?
한 페이지 추출은 성공했는데, 왜 그룹추출은 실패할까?
로그인 후에만 접근할 수 있는 페이지의 추출 설정을 알아봅니다.

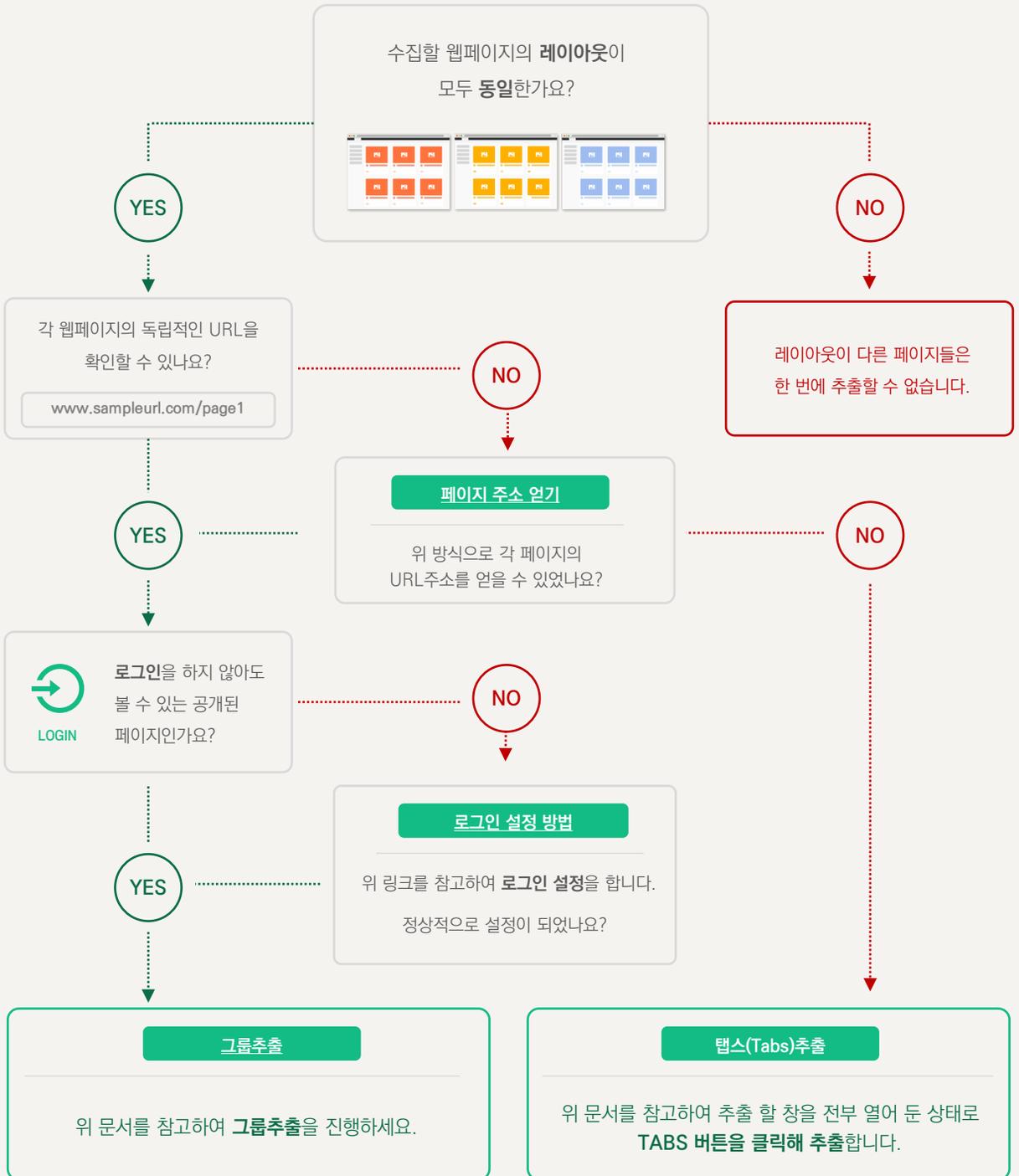


웹사이트에서 차단당한 경우

설마..나..차단 당한걸까?
반복작업으로 인해 웹사이트가 IP를 차단했을 경우
오류 화면과 해결 방법을 알아봅니다.

여러 페이지 수집 시 추출방식 선택

여러 페이지를 한 번에 추출하는 방법으로는 크게 그룹추출과 탭스가 있습니다. 사용자의 상황에 맞춰 추출방식을 선택합니다. 아래의 체크리스트는 추출방식 선택을 위한 1차 가이드입니다.

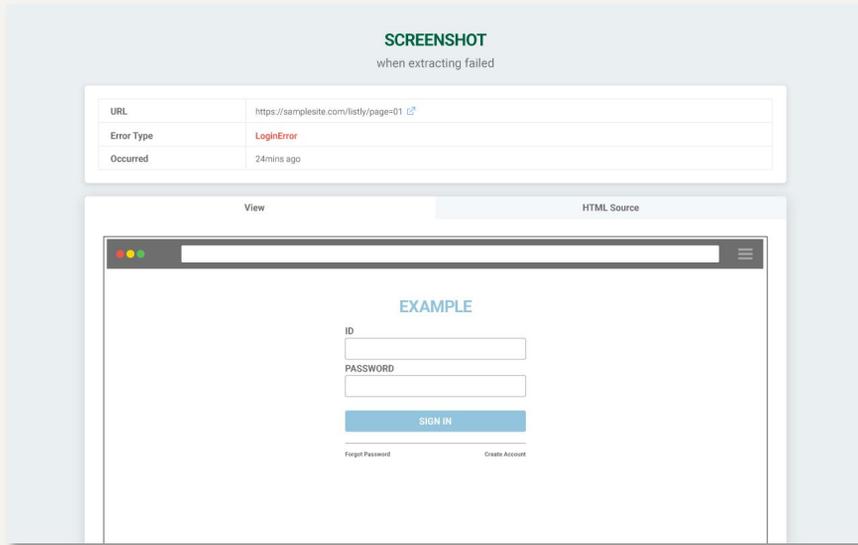


로그인 설정이 필요한 경우

리스틀리는 사용자의 쿠키를 사용해 로그인 후에만 접근이 가능한 웹페이지도 확장프로그램을 통한 1page->1excel 추출이 가능합니다. 하지만 쿠키는 임시파일로 일정 시간 경과 후에는 정보가 사라지기 때문에 아래의 경우 별도의 로그인 설정이 필요합니다.

1. 스케줄을 설정하여 일정 시간에 자동 재추출되는 태스크
2. URL입력만으로 여러 페이지를 추출하는 그룹 태스크

Ex. 로그인 오류 화면



오류 화면에는 로그인 화면, 웹사이트의 메인 화면, 타깃 페이지의 바로 전 단계 화면 등이 표시됩니다.

로그인 설정 방법

단, 아래의 상황은 beta기능으로 해결이 어렵습니다.

1. 2단계 로그인, 해외접속 금지 등의 설정으로 보안이 강력한 웹사이트(메일 서비스를 제공하는 웹사이트 대부분)
2. 아이디를 입력한 후 다음페이지로 이동해 비밀번호를 입력하는 UI
3. 로그인 창이 모달 팝업창으로 등장해 별도의 URL을 확인할 수 없는 웹사이트
4. 로그인 페이지의 URL이 임의문자열을 포함(복잡한 URL주소)하고 있는 웹사이트

로그인 설정이 원활하지 않은 경우 아래의 방법으로 데이터를 추출하세요.

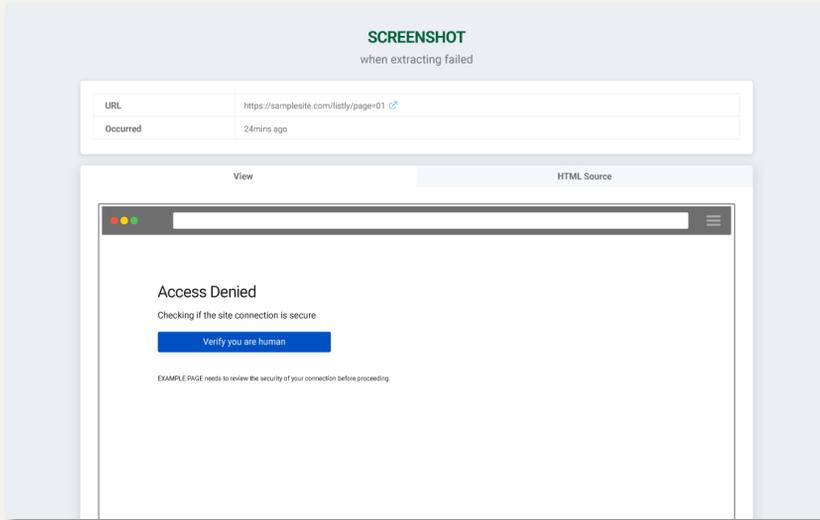
TABS

로그인 한 상태에서 추출할 웹페이지를 하나의 브라우저 창 안에서 각 탭으로 열어두고 데이터를 추출합니다. 추출작업을 마친 데이터는 자동으로 한 폴더로 합쳐지며, 하나의 엑셀로 내보낼 수 있습니다.(폴더 다운로드) 단, 해당 작업은 재추출이 불가능합니다.

웹사이트에서 차단당한 경우

데이터 수집 과정 중 한 웹사이트에서 단시간에 많은 반복작업을 수행할 경우 해당 웹사이트가 IP 접근을 차단 하는 경우가 있습니다. 이와 같은 상황은 프록시 설정으로 해결할 수 있습니다.

Ex. 사이트 접속 차단 오류 화면

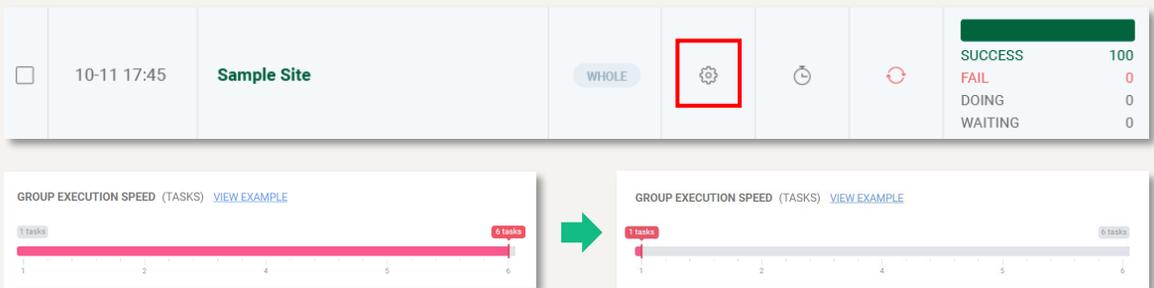


접속 차단으로 인한 오류 시 정상접근이 아님을 나타내는 예러 화면이 나타납니다.

프록시 설정 방법

+) 추출속도 조절

웹사이트에서 IP를 차단하는 상황은 '단시간'에 '많은 양의 반복작업'이 이루어질 때 입니다. 추출 환경을 더 안정적으로 만들고 싶다면 **그룹설정** → **그룹 추출 속도**를 1로 낮춰보세요.



특정 그룹이 아닌 **전체 태스크의 추출 속도**를 조절하고 싶을 때에는 데이터보드 우측 상단의 실행속도를 조절하세요.

