# 시나리오:
# 네이버 기사 크롤링 - UI

Power Automate Desktop 활용 자동화 역량 강화 시나리오

FDX NETWORKS

문의:
rpakr.com/kakao

본 웹페이지 외의 공간에서 저작물 공유 또는 게시를 금지합니다.

# 자료 내 약속 사항

- **변수명 작성**

| 표기 | 변수형 | 예시 |
|---|---|---|
| int_ | 숫자 | int_AppleCount, int_Age,.... |
| str_ | 문자열 | str_Name, str_filename,.... |
| date_ | 날짜 | date_Current, date_Yesterday,... |
| list_ | 목록 | list_Book,... |
| dt_ | 데이터테이블 | dt_Chess,... |
| v_ | 위의 경우에 해당하지 않는 변수 | vPDF파일,... |
| (없음) | 작업액션에서 기본 설정되는 변수명 그대로 | LoopIndex, ... |

- **실습 폴더 설정**

  바탕화면\RPA_MSPAD 폴더가 존재하고 있다는 가정하에 진행됩니다.

본 웹페이지 외의 공간에서 저작물 공유 또는 게시를 금지합니다.

##

# 시나리오: 네이버 기사 크롤링 – UI

**[실습 내용]**

1. 네이버 뉴스 기사 리스트의 기사 제목, 기사 링크, 언론사를 크롤링합니다.
2. 정해진 템플릿 시트에 크롤링 된 기사를 정리합니다.
3. 키워드.xlsx 파일에서 키워드를 읽어 키워드별로 기사를 정리합니다.

▼ 뉴스 템플릿

| 제목 | 원본링크 | 언론사 |
|---|---|---|
| | | |

▼ 키워드 예시

| 키워드 |
|---|
| 환율 |
| IT |
| 서울 |



본 웹페이지 외의 공간에서 저작물 공유 또는 게시를 금지합니다.

# 시나리오: 네이버 기사 크롤링 – UI

## [실습 전 준비 세팅]

- 실습에서 chrome 브라우저를 사용합니다.
  확장 프로그램이 설치되어 있어야 합니다.
  *rpakr.com/chrome 확인
- 네이버 뉴스 탭을 활용합니다. 네이버에 접속 가능해야 합니다.
- ‘뉴스키워드.xlsx’ 파일과 ‘뉴스크롤링템플릿_UI.xlsx’ 파일을 바탕화면\RPA_MSPAD 폴더에 배치하세요.

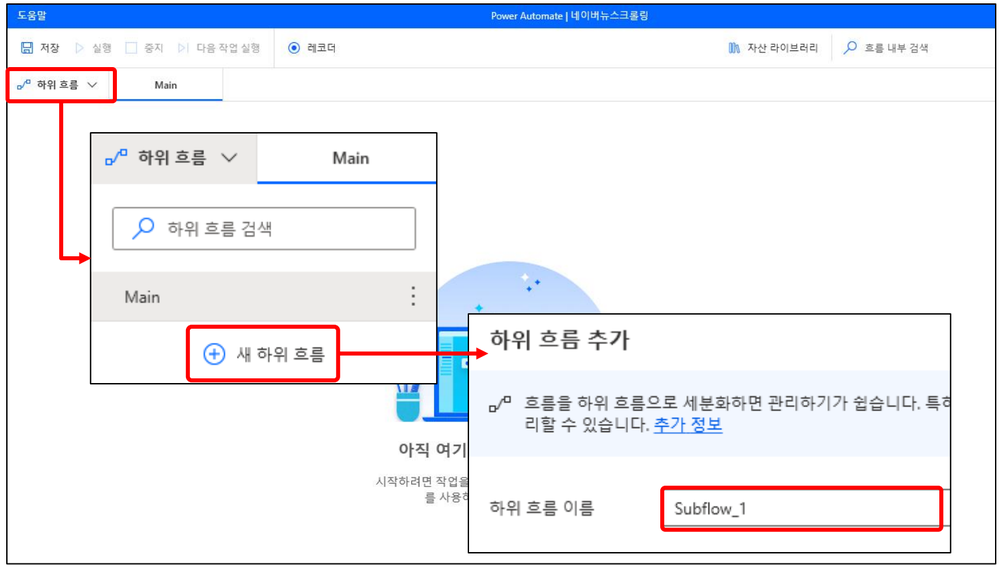
## [실습 주의사항]

- 해당  자료는 데이터테이블을 사용하여 진행합니다. 관련 개념을 익히고 진행하는 것을 추천합니다.



본 웹페이지 외의 공간에서 저작물 공유 또는 게시를 금지합니다.

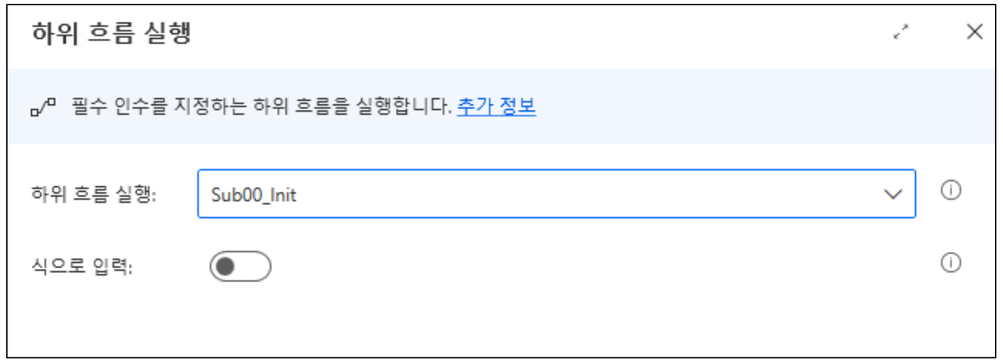
# 시나리오: 네이버 기사 크롤링 – UI




**01** 가장 먼저, 시나리오 구현을 위한 틀을 제작해야 합니다.
원하는 진행 사항에 맞추어 하위 흐름을 생성합니다.

이번 시나리오에서는 4개의 하위 흐름을 생성합니다.

| 하위흐름명 | 흐름 설명 |
|---|---|
| Sub00_Init | 시나리오에서 사용하는 전반적인 변수를 초기화합니다. |
| Sub01_GetKeyword | 엑셀 파일에서 모든 키워드를 읽습니다. |
| Sub02_GetNews | 키워드에 따라 뉴스를 크롤링합니다. |
| Sub03_WriteExcel | 엑셀에 작성합니다. |

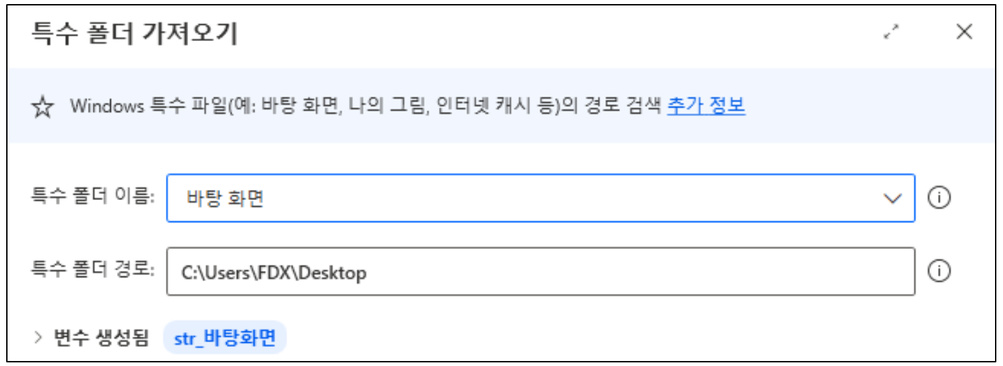



**02** [하위 흐름 시행] 작업액션으로 하위 흐름을 호출하여 실행시킬 수 있습니다.

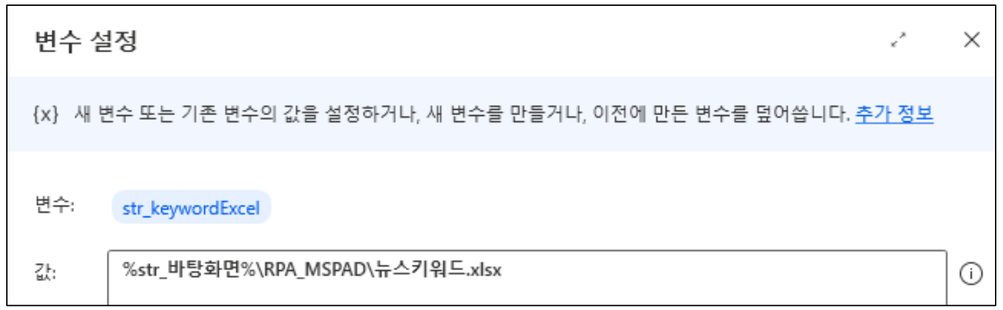
Sub00_Init과 Sub01_GetKeyword부터 배치하겠습니다.

본 웹페이지 외의 공간에서 저작물 공유 또는 게시를 금지합니다.

# 시나리오: 네이버 기사 크롤링 – UI

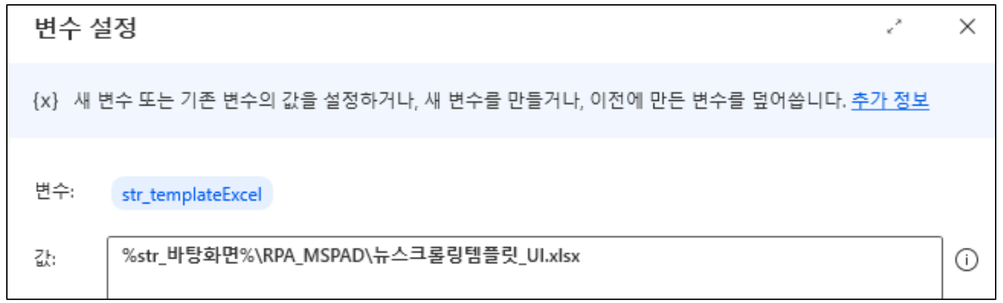



**03** Sub00_Init 하위흐름부터 시작하겠습니다.
[특수 폴더 가져오기] 작업 액션으로
바탕화면 경로를 설정합니다.




**04** [변수 설정]으로 검색할 키워드가 담긴 엑셀 파일의 경로명 및 파일명,
파일확장자까지 설정합니다.

변수: str_keywordExcel
값: %str_바탕화면%\RPA_MSPAD\뉴스키워드.xlsx



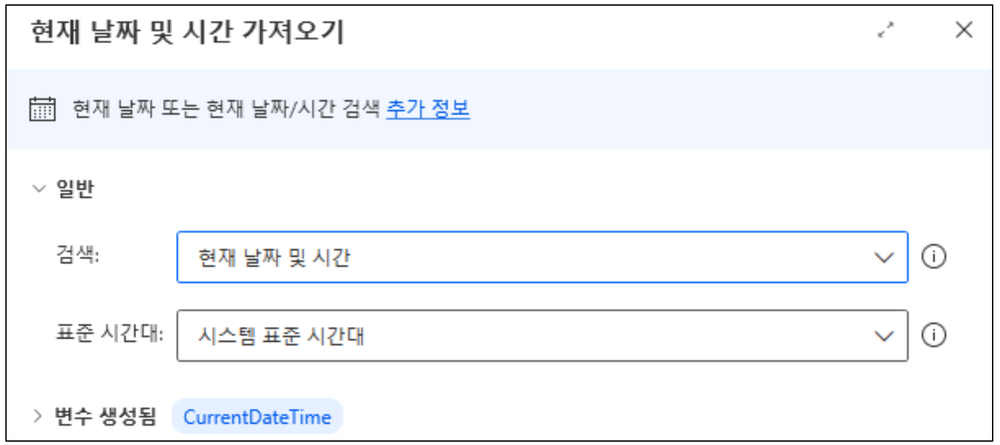

**05** [변수 설정]으로 크롤링한 데이터를 정리할 템프릿 엑셀 파일의 경로명
및 파일명, 파일확장자까지 설정합니다.

변수: str_keywordExcel
값: %str_바탕화면%\RPA_MSPAD\뉴스크롤링템플릿_UI.xlsx



본 웹페이지 외의 공간에서 저작물 공유 또는 게시를 금지합니다.

# 시나리오: 네이버 기사 크롤링 – UI




**06** [현재 날짜 및 시간 가져오기] 작업액션을 사용하여 오늘 날짜를 가져옵니다.

**07** 6번에서 가져온 오늘 날짜를 원하는 형태로 가공하기 위해 [텍스트로 날짜/시간 변환] 작업액션을 사용합니다.

변환할 날짜/시간: %CurrentDateTime%
사용자 지정 형식: yyMMdd
변수명: str_CurrentDate



본 웹페이지 외의 공간에서 저작물 공유 또는 게시를 금지합니다.

# 시나리오: 네이버 기사 크롤링 – UI
## Sub00_Init 전체 코드

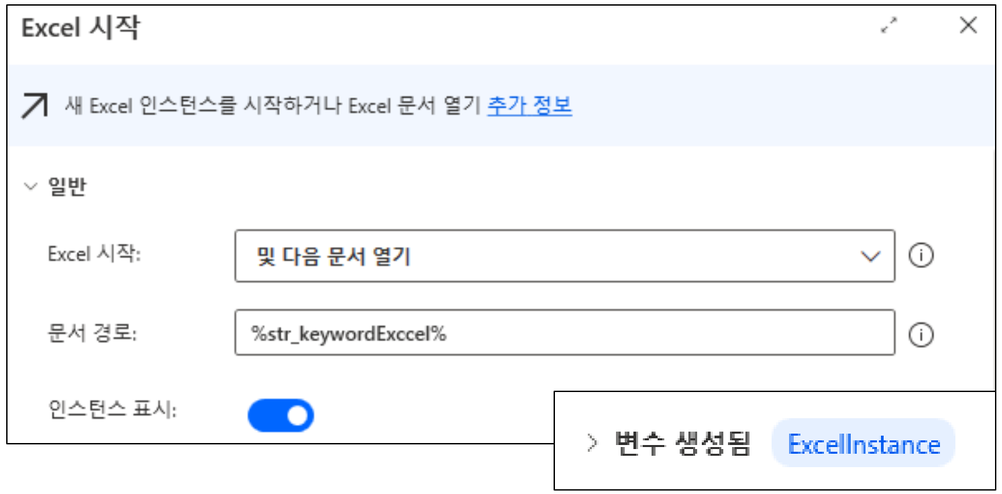
◀ 폴더/파일 설정

◀ 날짜 설정



본 웹페이지 외의 공간에서 저작물 공유 또는 게시를 금지합니다.

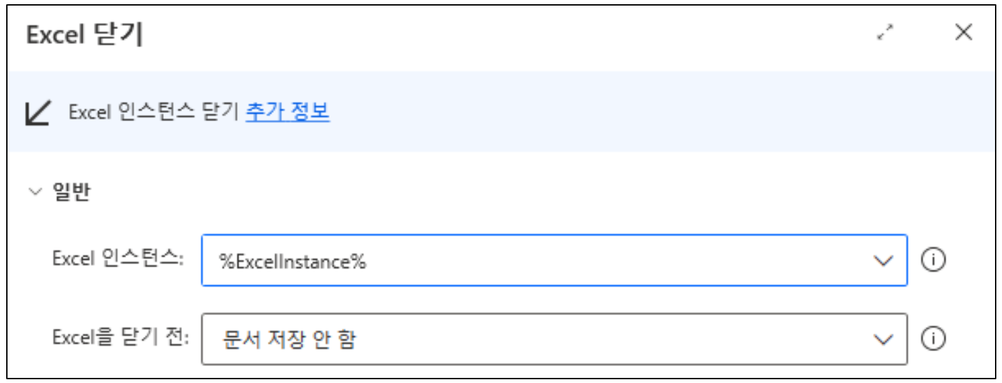
# 시나리오: 네이버 기사 크롤링 – UI




**08** 다음으로 진행할 하위흐름은 Sub01_GetKeyword입니다.
해당 하위 흐름은 엑셀 파일에서 검색 키워드를 모두 가져옵니다.

먼저, [Excel 시작] 작업액션을 사용하여 앞서 설정한 변수로 파일을 시작합니다.

문서 경로: %str_keywordExcel%

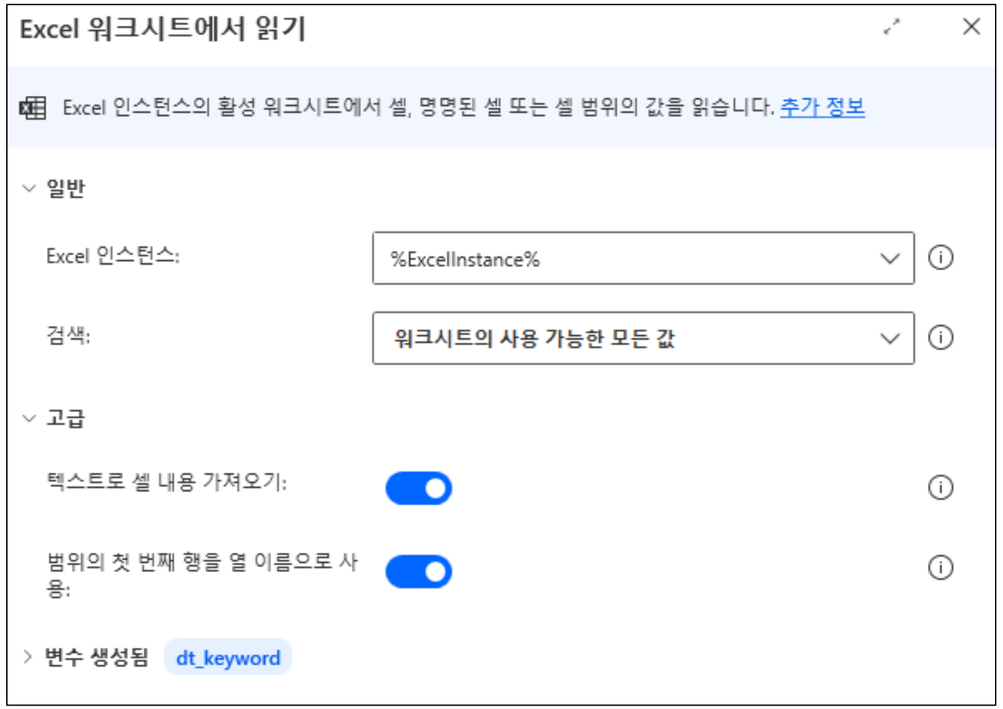



**09** 마지막으로 해당 파일을 닫을 때는 [Excel 닫기] 작업액션을 사용합니다.



본 웹페이지 외의 공간에서 저작물 공유 또는 게시를 금지합니다.

# 시나리오: 네이버 기사 크롤링 – UI




**10** [Excel 워크시트에서 읽기] 작업액션과 '워크시트의 사용 가능한 모든 값' 옵션을 사용하여 엑셀 시트의 값을 읽을 수 있습니다.

키워드가 담긴 엑셀 파일을 보면, 첫 번째 행은 그저 헤더로서 검색 시에는 필요 없는 값입니다.
'범위의 첫 번째 행을 열 이름으로 사용' 옵션을 활성화합니다.

변수: dt_keyword

▼ 엑셀

| 키워드 |
| --- |
| 환율 |
| IT |
| 서울 |

▼ PAD 데이터테이블

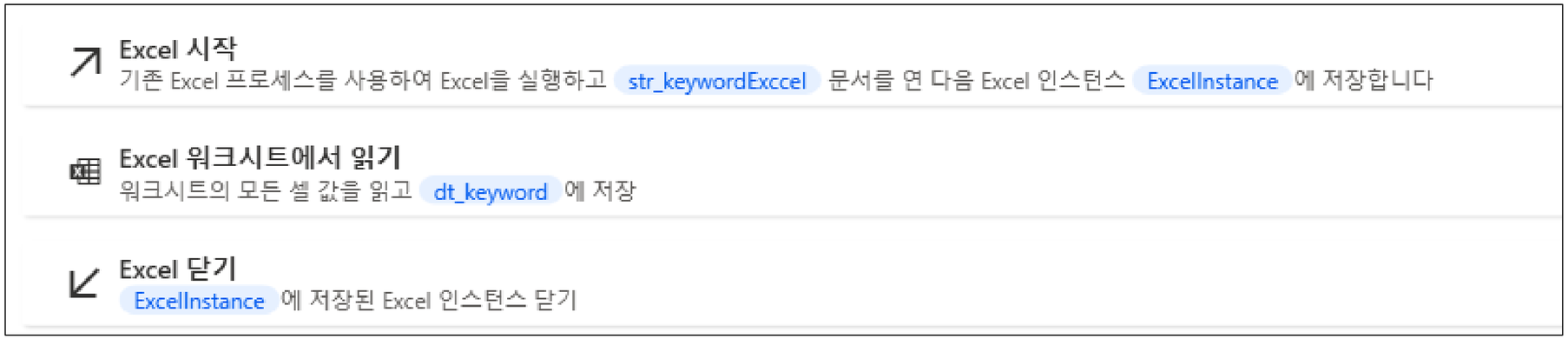
dt_keyword (데이터 테이블)

| | 키워드 |
| --- | --- |
| 0 | 환율 |
| 1 | IT |
| 2 | 서울 |



본 웹페이지 외의 공간에서 저작물 공유 또는 게시를 금지합니다.

# 시나리오: 네이버 기사 크롤링 – UI
## Sub01_GetKeyword 전체 코드




◀ 결과

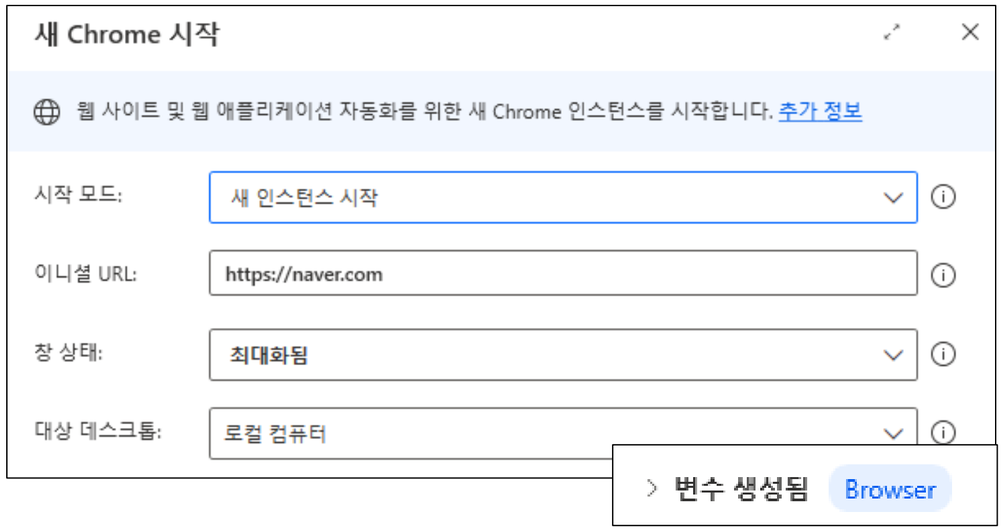
dt_keyword (데이터 테이블)

| | 키워드 | |
|---|---|---|
| 0 | 환율 | |
| 1 | IT | |
| 2 | 서울 | |



본 웹페이지 외의 공간에서 저작물 공유 또는 게시를 금지합니다.

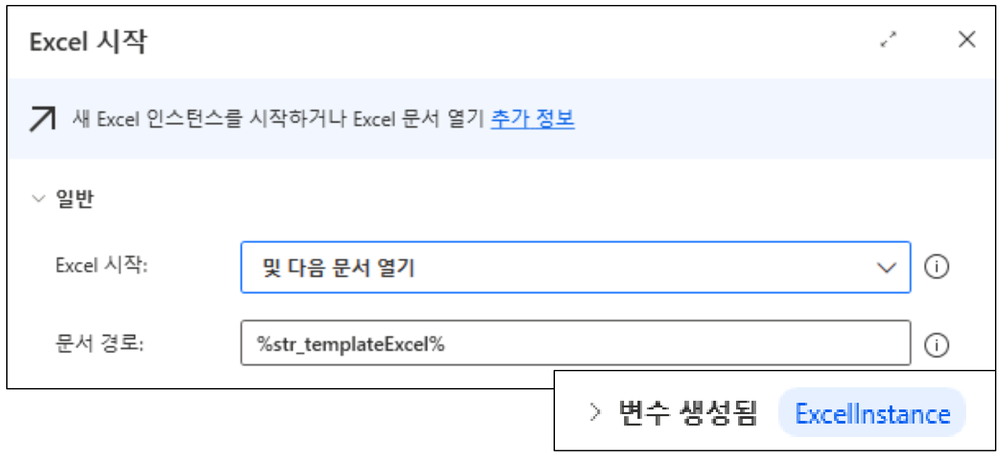
# 시나리오: 네이버 기사 크롤링 – UI




**11** 다시 Main으로 돌아와 진행합니다.

네이버에서 크롤링하기 위해 [새 Chrome 시작] 작업액션을 사용해 브라우저를 시작합니다.

URL: https://naver.com



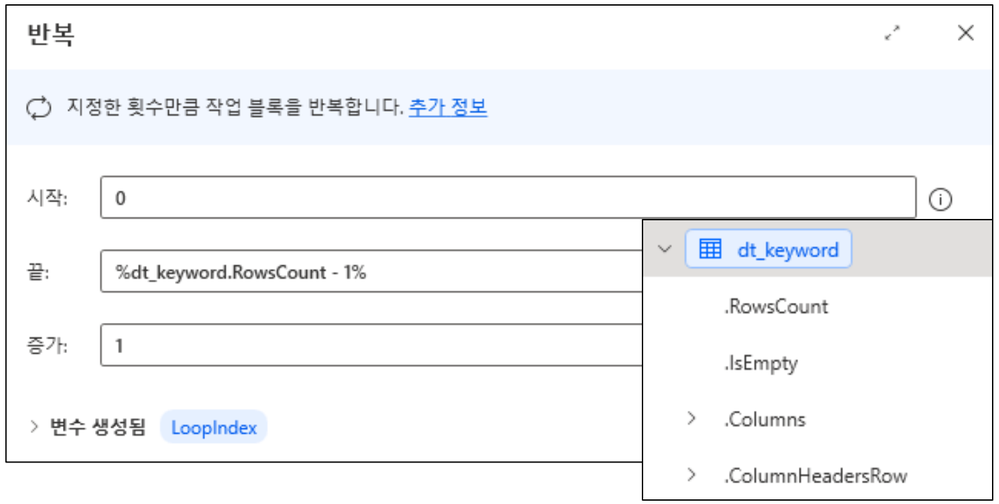

**12** [Excel 시작] 작업액션을 사용하여 템플릿 엑셀 파일을 시작합니다. 해당 파일의 경로는 앞서 변수로 설정했습니다. (Sub00_Init)

문서 경로: %str_templateExcel%



본 웹페이지 외의 공간에서 저작물 공유 또는 게시를 금지합니다.

# 시나리오: 네이버 기사 크롤링 – UI




**13** [반복] 작업액션을 사용해 키워드만큼 웹 크롤링 작업과 엑셀 작성 작업이 계속 반복되도록 합니다.

데이터테이블의 인덱스는 0부터 시작하며 행 개수는 .RowsCount 속성을 사용하여 유동적으로 알 수 있습니다.

시작: 0
끝: %dt_keyword.RowsCount - 1%
증가: 1

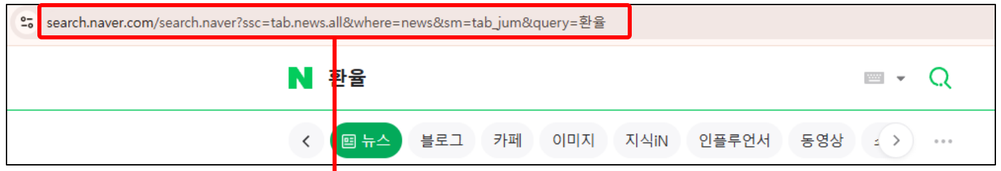
**14** 반복문 내 Sub02와 Sub03 하위흐름 실행을 배치합니다.



본 웹페이지 외의 공간에서 저작물 공유 또는 게시를 금지합니다.

# 시나리오: 네이버 기사 크롤링 – UI

**15** Sub02_GetNews 하위 흐름입니다. 해당 하위흐름을 사용하여 검색된 뉴스 페이지로 이동합니다.




https://search.naver.com/search.naver?ssc=tab.news.all&where=news&sm=tab_jum&query=환율 ◀ 검색어

https://search.naver.com/search.naver?ssc=tab.news.all&where=news&sm=tab_jum&query=%dt_keyword[LoopIndex][0]%

## TIP!

**링크 내 '%'가 포함되어 있는데 복사해서 URL에 붙여 넣을 경우 오류가 떠요!**

▷ Power Autoamte Desktop은 %를 변수 선언 문자로 활용하기 때문입니다.
'%%' 이렇게 연달아 쓰면 오류가 해결되고, %로 인식됩니다.

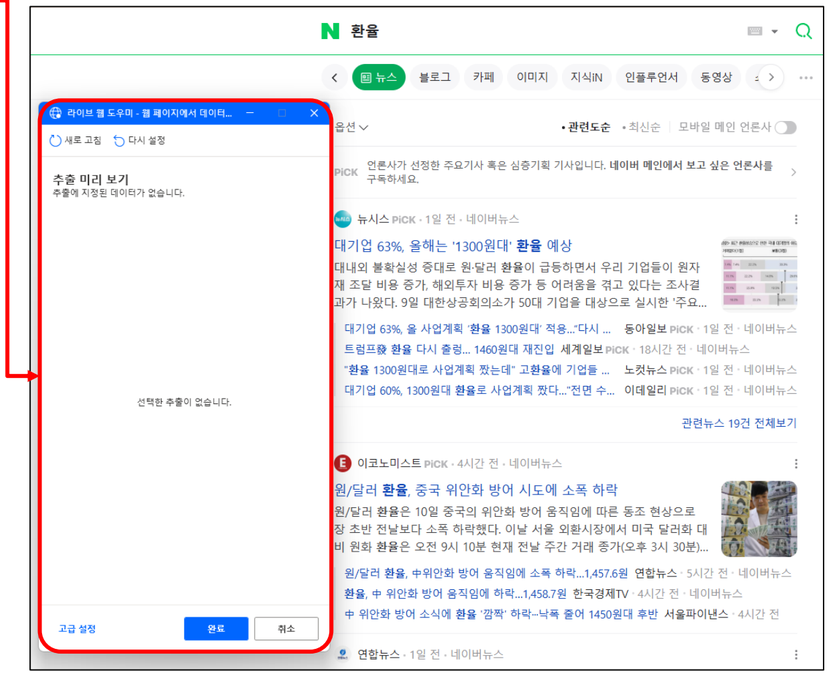
본 웹페이지 외의 공간에서 저작물 공유 또는 게시를 금지합니다.

# 시나리오: 네이버 기사 크롤링 – UI

## 16-(1)

웹 크롤링을 위해 [웹 페이지에서 데이터 추출] 작업액션을 사용합니다.
해당 작업액션을 그대로 PAD에 띄운 후, 웹으로 바로 이동합니다.

바로 라이브 웹 도우미가 뜨는 것을 확인할 수 있습니다.






본 웹페이지 외의 공간에서 저작물 공유 또는 게시를 금지합니다.

# 시나리오: 네이버 기사 크롤링 – UI

## 16-(2)

먼저, 기사 제목에 마우스를 올린 다음, 오른쪽 클릭합니다.
요소 값 추출을 선택하면, 어떤 요소 값을 선택할 것인지 나타납니다.

먼저 텍스트를 선택한 뒤, 다시 진행하여 Href도 추출합니다.

## 16-(3)

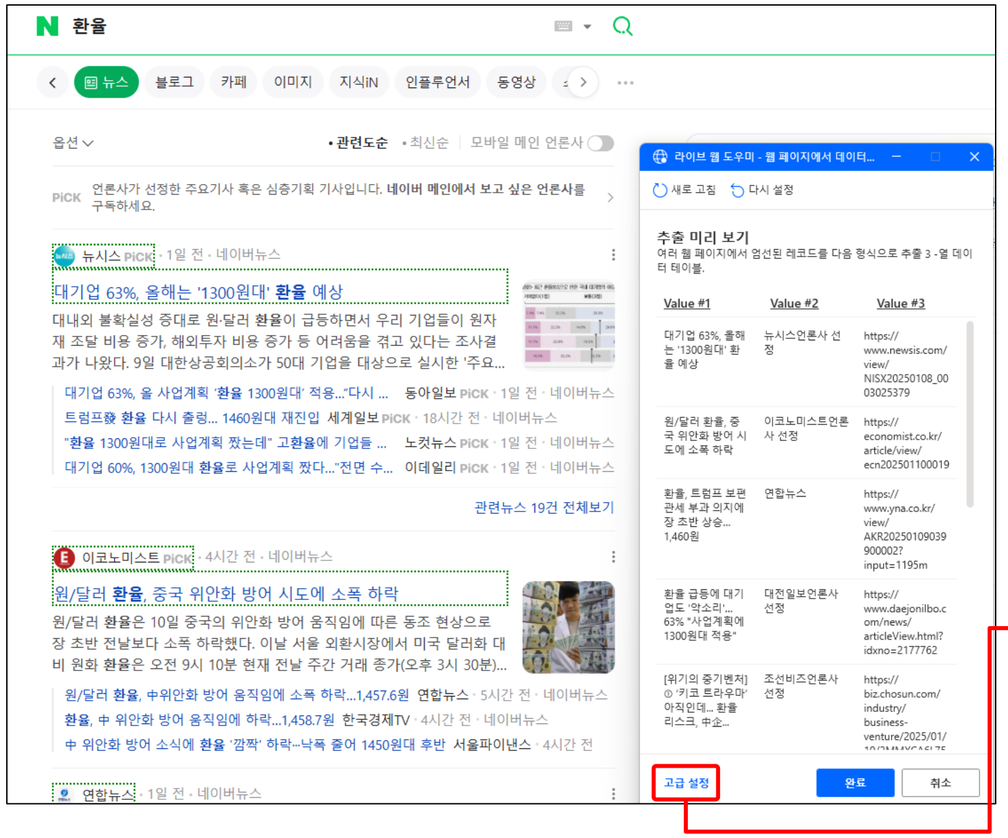
언론사에 마우스를 올린 뒤, 텍스트를 추출합니다.

◀ 기사 하나 추출 완료 시 상태



본 웹페이지 외의 공간에서 저작물 공유 또는 게시를 금지합니다.

# 시나리오: 네이버 기사 크롤링 – UI

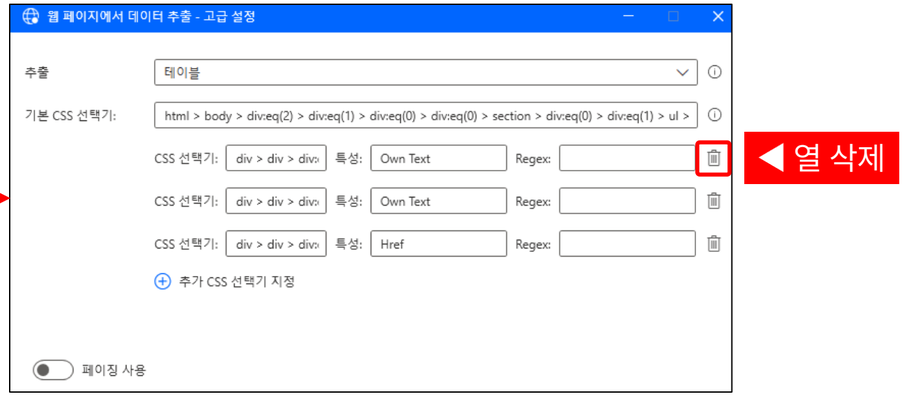



## 16-(4)

두 번째와 세 번째 기사를 하나씩 진행하다 보면 PAD가 자동으로 인식하여 전체 데이터를 크롤링합니다.

완료 버튼을 눌러 설정을 끝냅니다.

## TIP!



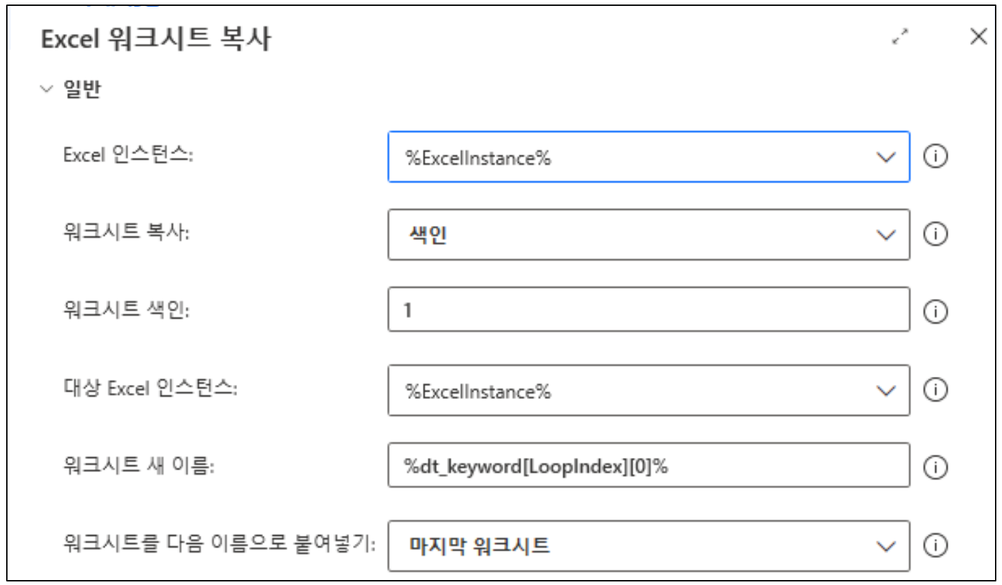

◀ 열 삭제

* 해당 창을 통해 원하지 않는 데이터를 삭제, 혹은 다른 데이터를 추출하도록 편집할 수 있습니다.



본 웹페이지 외의 공간에서 저작물 공유 또는 게시를 금지합니다.

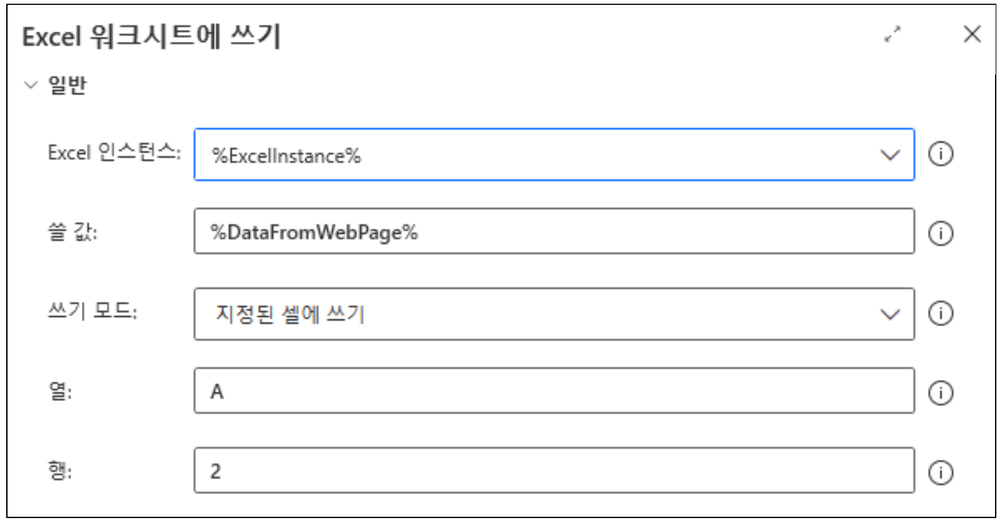
# 시나리오: 네이버 기사 크롤링 – UI




**17** Sub03_WriteExcel 하위 흐름입니다.

[Excel 워크시트 복사] 작업액션을 활용하여 기존의 템플릿을 복사해 새로 생성합니다.

워크시트 색인: 1
새 이름: %dt_keyword[LoopIndex][0]%
워크시트를 다음 이름으로 붙여넣기: 마지막 워크시트




**18** [Excel 워크시트에 쓰기] 작업액션을 통해 크롤링한 데이터를 엑셀에 작성합니다.

쓸 값: %DataFromWebPage%
열: A
행: 2

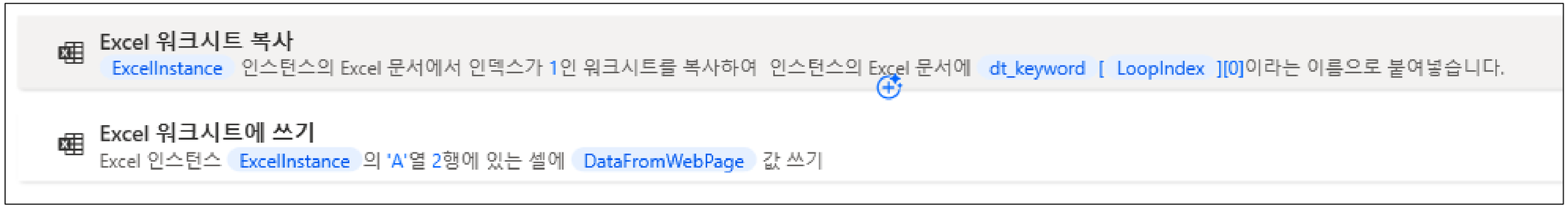


본 웹페이지 외의 공간에서 저작물 공유 또는 게시를 금지합니다.

# 시나리오: 네이버 기사 크롤링 – UI

## Sub02_GetNews 전체 코드

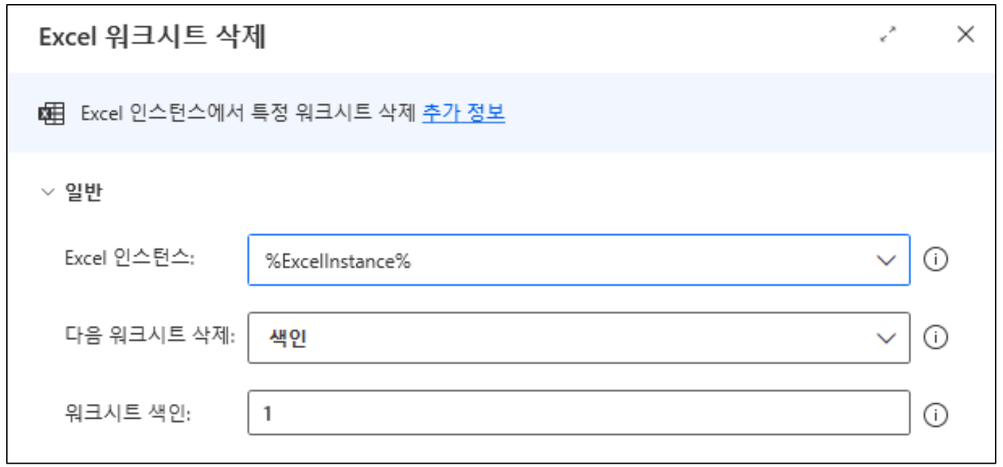
## Sub03_WriteExcel 전체 코드

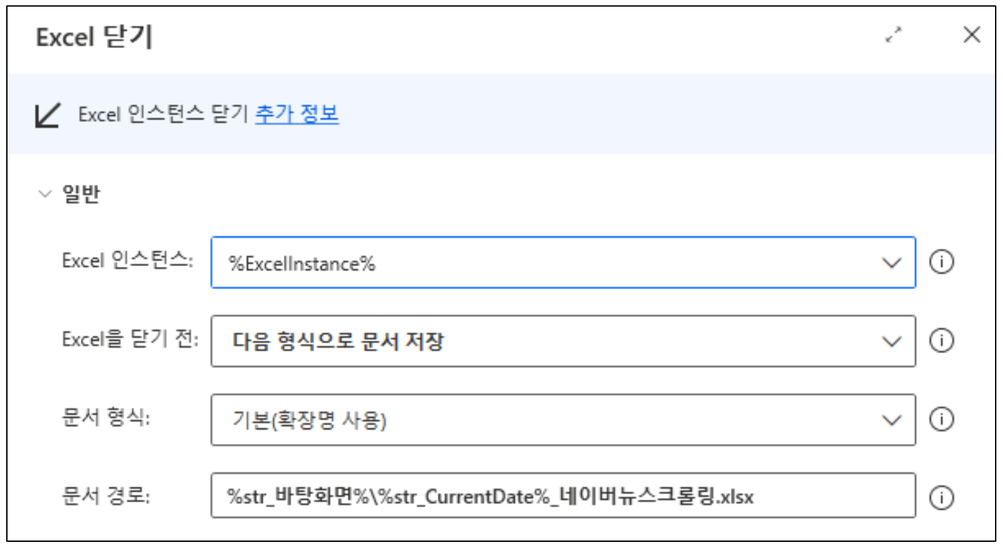





본 웹페이지 외의 공간에서 저작물 공유 또는 게시를 금지합니다.

# 시나리오: 네이버 기사 크롤링 – UI




**19**

Main 작업입니다.

앞 순서까지 완료되었다면, 가장 처음에 템플릿 시트 그리고 각 검색어 시트가 존재합니다.
[Excel 워크시트 삭제] 작업액션으로 첫 번째 시트를 삭제합니다.

워크시트 색인: 1

**20**

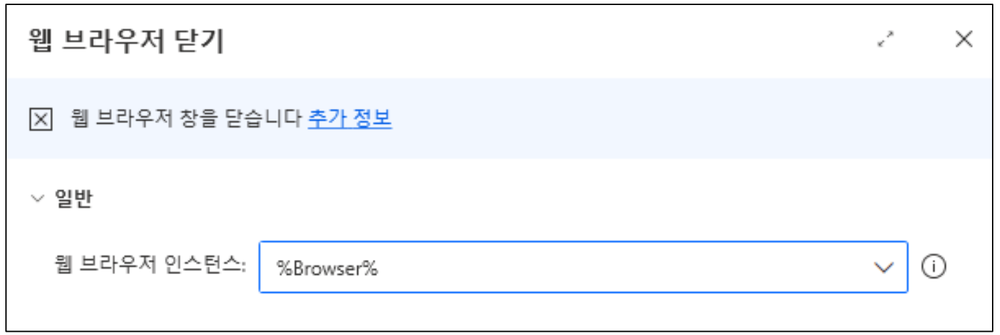
[Excel 닫기] 작업액션을 사용해서 결과 파일을 다른 이름 저장합니다.

문서경로:
%str_바탕화면%\%str_CurrentDate%_네이버뉴스크롤링.xlsx



본 웹페이지 외의 공간에서 저작물 공유 또는 게시를 금지합니다.

# 시나리오: 네이버 기사 크롤링 – UI




**21** 웹 브라우저까지 종료합니다.



본 웹페이지 외의 공간에서 저작물 공유 또는 게시를 금지합니다.

# 시나리오: 네이버 기사 크롤링 – UI
## Main 전체 코드

본 웹페이지 외의 공간에서 저작물 공유 또는 게시를 금지합니다.