

# 시나리오: PDF 데이터 추출

---

Power Automate Desktop 활용 자동화 역량 강화 시나리오

# 자료 내 약속 사항

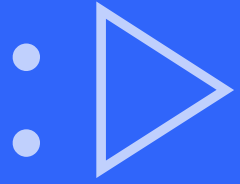
■ 변수명 작성

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.

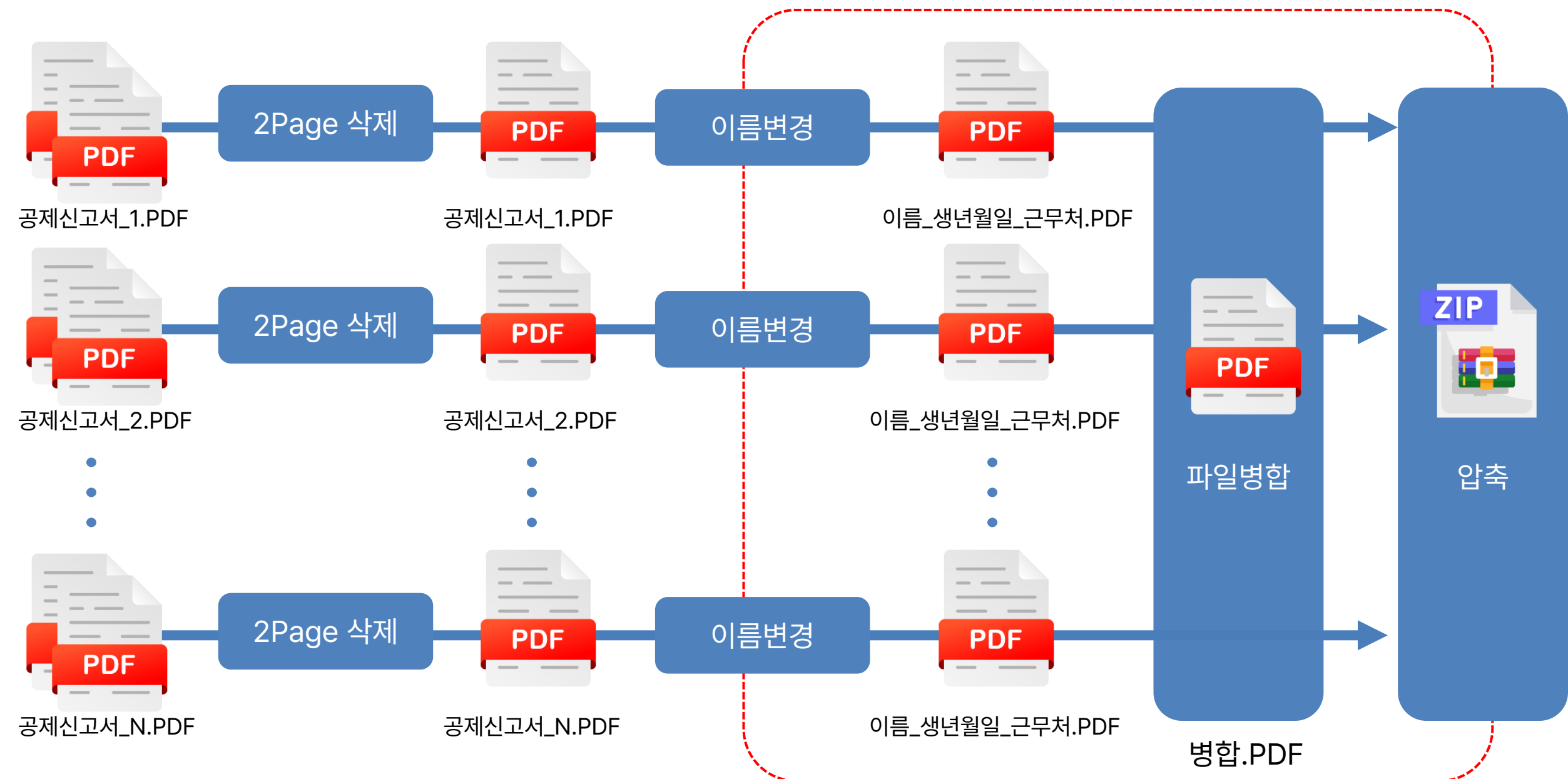




# 시나리오: PDF 데이터 추출

## [실습 내용]

1. PDF에서 원하는 데이터를 추출하여 파일명을 변경합니다.
2. 모든 파일에 대하여 적용합니다.
3. PDF 파일 병합 후 압축합니다.



# 시나리오: PDF 데이터 추출

## [실습 전 준비 세팅]

- [rpakr.com/data/RPA실습실습\\_공제신고서.zip](http://rpakr.com/data/RPA실습실습_공제신고서.zip) 링크를 통해 이번 시나리오에서 사용할 파일을 다운로드할 수 있습니다.
- 다운로드 받은 폴더를 압축 해제하고 바탕화면\RPA\_MSPAD\PDF데이터추출 폴더에 이동시킵니다.
- 바탕화면\RPA\_MSPAD\결과 폴더를 생성합니다.

## [실습 주의사항]

- 이번 실습은 작은 파트부터 시작해 반복문을 하나씩 적용합니다.

## 시나리오: PDF 데이터 추출

[illegible]

01

공제신고서의 PDF파일에서 가져올 영역의 데이터를 선정합니다.

이름, 생년월일, 근무지를 가져오도록 하겠습니다.

02

[특수 폴더 가져오기] 작업 액션으로  
바탕화면 경로를 설정합니다.

# 시나리오: PDF 데이터 추출

PDF에서 텍스트 추출

PDF 파일에서 텍스트 추출 [추가 정보](#)

일반

PDF 파일: %str\_바탕화면%\RPA\_MSPAD\PDF데이터추출\공제신고서\_1.pdf

추출할 페이지: 모두

고급

변수 생성됨 str\_PDF추출텍스트

03 [PDF에서 텍스트 추출] 작업액션을 사용합니다. 해당 작업액션은 'readable(PDF를 열었을 때 드래그로 긁히는)' 텍스트를 모두 가져올 수 있습니다.

PDF 파일:

%str\_바탕화면%\RPA\_MSPAD\PDF데이터추출\공제신고서\_1.pdf

변수명: str\_PDF추출텍스트

변수 값

str\_PDF추출텍스트 (텍스트 값)

1 ■ 소득세법 시행규칙 [별지 제37호서식(1)] <개정 2021. 10. 28.> (9쪽 중 제1쪽)

2 소득·세액 공제신고서/근로소득자 소득·세액 공제신고서(년 소득에 대한 연말정산용)

3 ※ 근로소득자는 신고서에소득·세액 공제 증명서류를 첨부하여원천징수의무자(소속 회사 등)에게 제출하며, 원천징수의무자는

4 인하여 근로소득 세액계산을 하고 근로소득자에게 즉시 근로소득원천징수영수증을발급해야 합니다. 연말정산 시 근로소득자에게

5 경우 원천징수의무자는 근로소득자에게 환급세액을 지급해야 합니다.

6 소득자 성명 강현우 주민등록번호 870507-2468135

7 근무처 명칭 KT 사업자등록번호 - -

8 세대주 여부 [ ]세대주 [ ]세대원 국 적 (국적 코드:)

9 근무기간 ~ 감면기간 ~

10 거주구분 [ ]거주자 [ ]비거주자 거주지국 (거주지국 코드:)

11 인적공제 항목

12 변동 여부 [ ]전년과 동일 [ ]변동 분납신청여부 [ ]신청 [ ]미신청

13 원천징수세액 선택 [ ]120% [ ]100% [ ]80% ※ 근로소득자 본인이 원하는 경우 매월 원천징수하는 세액을

14 법령상 세액의 120%, 100%, 80% 중 선택할 수 있습니다.

15 I .

16 인

17 적

18 구

\* 추출하고자 하는 데이터 중 성명은 '성명'과 '주민등록번호' 사이에, 생년월일은 '주민등록번호'와 '-', 근무처는 '명칭'과 '사업자등록번호' 사이 있는 것을 확인할 수 있습니다.

# 시나리오: PDF 데이터 추출

텍스트 자르기

주어진 텍스트에서 지정된 텍스트 플래그 이전, 이후 또는 사이에 발생하는 텍스트 값을 검색합니다 [추가 정보](#)

일반

원본 텍스트: %str\_PDF추출텍스트%

모드: 지정된 두 플래그 사이에 텍스트 가져오기

시작 플래그: 소득자 성명

종료 플래그: 주민등록번호

대/소문자 무시: ☐

> 변수 생성됨 str\_성명

시작 플래그: 주민등록번호

종료 플래그: -

시작 플래그: 근무처 명칭

종료 플래그: 사업자등록번호

04

[텍스트 자르기] 작업액션으로 특정 문자열 사이에 있는 텍스트를 가져올 수 있습니다. 3번에서 확인한 것에 따라 성명, 생년월일, 근무처를 설정합니다.

시작 플래그: 소득자 성명  
종료 플래그: 주민등록번호  
변수명: str\_성명

시작 플래그: 주민등록번호  
종료 플래그: -  
변수명: str\_생년월일

시작 플래그: 근무처 명칭  
종료 플래그: 사업자등록번호  
변수명: str\_근무처

실행 결과 ▼

(x) str\_성명 강현우

(x) str\_생년월일 870507

(x) str\_근무처 KT

# 시나리오: PDF 데이터 추출

변수 설정

{x} 새 변수 또는 기존 변수의 값을 설정하거나, 새 변수를 만들거나, 이전에 만든 변수를 덮어씁니다. [추가 정보](#)

변수: str\_PDF결과

값: %str\_성명%\_%str\_생년월일%\_%str\_근무처%

텍스트 대체

다른 텍스트를 가진 지정된 하위 텍스트의 모든 항목을 대체합니다. 정규식과 함께 사용될 수도 있습니다

[추가 정보](#)

일반

구문 분석할 텍스트: %str\_PDF결과%

찾을 텍스트: %' '%

찾고 대체하기를 위해 정규식 사용: ☐

대/소문자 무시: ☐

바꿀 내용: %''%

이스케이프 시퀀스 활성화: ☐

변수 생성됨

str\_PDF결과

05 4번의 결과를 PDF 파일명으로 변경할 예정입니다.  
변경하기 전, 우선 [변수 설정]으로 제목을 변수에 저장하겠습니다.

변수명: str\_PDF결과  
값: %str\_성명%\_%str\_생년월일%\_%str\_근무처%

str\_PDF결과 (텍스트 값)

1 강현우\_870507\_KT

\* 실행 시,  
공백이 있는 것을 확인할 수 있습니다.

06 [텍스트 대체 작업액션으로 공백을 제거합니다.

찾을 텍스트: %' '%  
바꿀 내용: %''%


str\_PDF결과 (텍스트 값)

1 강현우\_870507\_KT




## 시나리오: PDF 데이터 추출

새 PDF 파일로 PDF 파일 추출


 PDF 파일에서 새 PDF 파일로 페이지 추출 [추가 정보](#)

▽ 일반


PDF 파일:



페이지 선택:




추출된 PDF 파일 경로:



파일이 있는 경우:

▽



> 고급

> 변수 생성됨 **v2페이지를삭제한PDF**

## 07 [새 PDF 파일로 PDF 파일 추출] 작업액션 사용시, 특정 페이지만 추출하며 파일명을 설정할 수 있습니다.

페이지 선택: 1

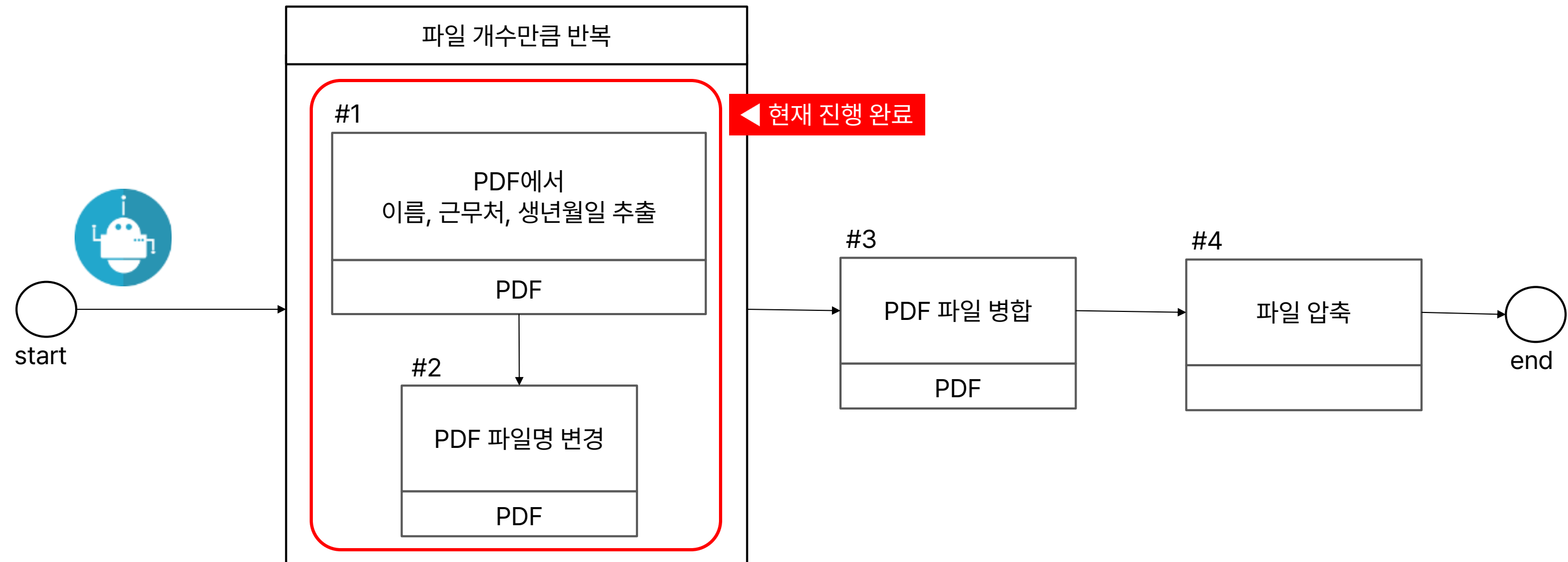
추출된 PDF 파일 경로:

%str\_바탕화면%\RPA\_MSPAD\결과\%str\_PDF결과%.pdf

## 현재 코드 상태 ▼

- ☆ 특수 폴더 가져오기  
바탕 화면 폴더의 경로를 가져오고 `str_바탕화면` 에 저장
- PDF에서 텍스트 추출  
`str_바탕화면 \RPA_MSPAD\PDF데이터추출\공제신고서_1.pdf` PDF 파일에서 `str_PDF추출텍스트` (으)로 텍스트 추출
- 텍스트 자르기  
텍스트 `str_PDF추출텍스트` 에서 지정된 플래그 '소득자 성명'와 '주민등록번호' 사이의 텍스트를 가져와서 `str_성명` 에 저장합니다.
- 텍스트 자르기  
텍스트 `str_PDF추출텍스트` 에서 지정된 플래그 '주민등록번호'와 '-' 사이의 텍스트를 가져와서 `str_생년월일` 에 저장합니다.
- 텍스트 자르기  
텍스트 `str_PDF추출텍스트` 에서 지정된 플래그 '근무처 명칭'와 '사업자등록번호' 사이의 텍스트를 가져와서 `str_근무처` 에 저장합니다.
- {x} 변수 설정  
변수 `str_PDF결과` 에 할당 값 `str_성명 '.' str_생년월일 '.' str_근무처`
- 텍스트 대체  
`str_PDF결과` 에서 '' 텍스트를 "(으)로 대체하고 결과를 `str_PDF결과` 에 저장
- 새 PDF 파일로 PDF 파일 추출  
`str_바탕화면 \RPA_MSPAD\PDF데이터추출\공제신고서_1.pdf` PDF 문서에서 `str_바탕화면 \RPA_MSPAD\결과\' str_PDF결과 '.pdf` PDF 파일로 1 페이지 추출 및 추출된 PDF를 `v2페이지를삭제한PDF` 에 저장

# 시나리오: PDF 데이터 추출



# 시나리오: PDF 데이터 추출

폴더의 파일 가져오기

폴더에서 파일 목록 검색 [추가 정보](#)

일반

폴더: %str\_바탕화면%\RPA\_MSPAD\PDF데이터추출

파일 필터: \*

하위 폴더 포함: ☐

고급

변수 생성됨 **Files**

08 이제 파일 하나가 아닌 모든 파일에 대하여 작업을 적용할 것입니다. 우선, [특수 폴더 가져오기] 작업액션 하단에 [폴더의 파일 가져오기] 작업액션을 사용하여 폴더의 모든 파일들을 가져옵니다.

변수 값

Files (목록 파일)

항목	
0	C:\Users\FDX\Desktop\RPA_MSPAD\PDF데이터추출\공제신고서_1.pdf
1	C:\Users\FDX\Desktop\RPA_MSPAD\PDF데이터추출\공제신고서_2.pdf
2	C:\Users\FDX\Desktop\RPA_MSPAD\PDF데이터추출\공제신고서_3.pdf

각각의 경우

목록, 데이터 테이블, 데이터 행의 항목을 반복하여 작업 블록을 반복적으로 실행합니다. [추가 정보](#)

반복할 값: %Files%

저장 위치: CurrentItem

09 파일 목록을 하나씩 반복하며 작업을 적용합니다. [각각의 경우] 작업액션을 사용합니다.

반복할 값: %Files%

# 시나리오: PDF 데이터 추출

PDF에서 텍스트 추출

PDF 파일에서 텍스트 추출 [추가 정보](#)

일반

PDF 파일:

새 PDF 파일로 PDF 파일 추출

PDF 파일에서 새 PDF 파일로 페이지 추출 [추가 정보](#)

일반

PDF 파일:

바탕 화면 > RPA_MSPAD > 결과	
이름	수정한 날짜
강동우_970822_LG전자.pdf	2025-01-09
강현우_870507_KT.pdf	2025-01-09
장미경_920210_SK텔레콤.pdf	2025-01-09

10 작업액션에서 본래 경로명과 파일명이 적혔던 부분을 %CurrentItem% 변수로 변경합니다.

현재 코드 상태 ▼

☆ 특수 폴더 가져오기  
바탕 화면 폴더의 경로를 가져오고 str\_바탕화면 에 저장

폴더의 파일 가져오기  
str\_바탕화면 \RPA\_MSPAD\PDF데이터추출 폴더에서 \*.\* 필터와 일치하는 파일을 검색하고 Files 에 저장

각각의 경우 Files 내 각 CurrentItem 의 경우

PDF에서 텍스트 추출  
CurrentItem PDF 파일에서 str\_PDF추출텍스트 (으)로 텍스트 추출

텍스트 자르기  
텍스트 str\_PDF추출텍스트 에서 지정된 플래그 '소독자 성명'와 '주민등록번호' 사이의 텍스트를 가져와서 str\_성명 에 저장합니다.

텍스트 자르기  
텍스트 str\_PDF추출텍스트 에서 지정된 플래그 '주민등록번호'와 '-' 사이의 텍스트를 가져와서 str\_생년월일 에 저장합니다.

텍스트 자르기  
텍스트 str\_PDF추출텍스트 에서 지정된 플래그 '근무처 명칭'와 '사업자등록번호' 사이의 텍스트를 가져와서 str\_근무처 에 저장합니다.

{x} 변수 설정  
변수 str\_PDF결과 에 할당 값 str\_성명 '\_' str\_생년월일 '\_' str\_근무처

텍스트 대체  
str\_PDF결과 에서 '' 텍스트를 "(으)로 대체하고 결과를 str\_PDF결과 에 저장

새 PDF 파일로 PDF 파일 추출  
CurrentItem PDF 문서에서 str\_바탕화면 \RPA\_MSPAD\결과\ str\_PDF결과 '.pdf' PDF 파일로 1 페이지 추출 및 추출된 PDF를 v2페이지를삭제한PDF 에 저장

End 끝

◀ 실행 결과

# 시나리오: PDF 데이터 추출

폴더의 파일 가져오기

폴더에서 파일 목록 검색 [추가 정보](#)

일반

폴더:

파일 필터:

하위 폴더 포함: ☐

PDF 파일 병합

여러 PDF 파일을 새로운 PDF 파일 하나로 병합 [추가 정보](#)

일반

PDF 파일:

병합된 PDF 경로:

파일이 있는 경우:

고급

변수 생성됨 MergedPDF

11 이번에는 결과 폴더의 모든 파일을 가져옵니다.  
해당 결과 폴더에는 반복문이 모두 끝난 후의, 1페이지씩 존재하는 pdf  
파일들이 있습니다.

변수 값	
Files (목록 파일)	
항목	
0	C:\Users\FDX\Desktop\RPA_MSPAD\결과\강동우_970822_LG전자.pdf
1	C:\Users\FDX\Desktop\RPA_MSPAD\결과\강현우_870507_KT.pdf
2	C:\Users\FDX\Desktop\RPA_MSPAD\결과\장미경_920210_SK텔레콤.pdf

12 앞서 구한 폴더의 pdf들을 하나의 pdf로 병합합니다.

pdf 파일: **%Files%**  
병합된 PDF 경로: **%str\_바탕화면%\RPA\_MSPAD\결과\병합.pdf**

## 시나리오: PDF 데이터 추출

ZIP 파일

ZIP 압축 파일로 하나 이상의 파일 또는 폴더 압축 [추가 정보](#)

일반

보관 파일 경로: 

%str\_바탕화면%\RPA\_MSPAD\PDF데이터추출결과

파일 ▼

📁

📄

{x}

압축할 파일: 

%str\_바탕화면%\RPA\_MSPAD\결과

폴더 ▲

압축 수준: 속도와의 최적의 조화

암호: 

🔒

▼

보관 파일 설명:

> 변수 생성됨

ZipFile

13

마지막으로 [ZIP 파일] 작업액션을 활용하여 결과 폴더를 압축합니다.  
해당 작업액은 하나 혹은 다수의 폴더 또는 파일을 설정하여 압축할 수 있습니다.

보관 파일 경로:

%str\_바탕화면%\RPA\_MSPAD\PDF데이터추출결과.zip

압축할 파일: %str\_바탕화면%\RPA\_MSPAD\결과



# 시나리오: PDF 데이터 추출

결과 코드 ▼

☆ 특수 폴더 가져오기
바탕 화면 폴더의 경로를 가져오고 str_바탕화면 에 저장
폴더의 파일 가져오기
str_바탕화면 \RPA_MSPAD\PDF데이터추출' 폴더에서 '*' 필터와 일치하는 파일을 검색하고 Files 에 저장
각각의 경우 Files 내 각 CurrentItem 의 경우
PDF에서 텍스트 추출
CurrentItem PDF 파일에서 str_PDF추출텍스트 (으)로 텍스트 추출
텍스트 자르기
텍스트 str_PDF추출텍스트 에서 지정된 플래그 '소득자 성명'와 '주민등록번호' 사이의 텍스트를 가져와서 str_성명 에 저장합니다.
텍스트 자르기
텍스트 str_PDF추출텍스트 에서 지정된 플래그 '주민등록번호'와 '.' 사이의 텍스트를 가져와서 str_생년월일 에 저장합니다.
텍스트 자르기
텍스트 str_PDF추출텍스트 에서 지정된 플래그 '근무처 명칭'와 '사업자등록번호' 사이의 텍스트를 가져와서 str_근무처 에 저장합니다.
변수 설정
{x} 변수 str_PDF결과 에 할당 값 str_성명 '.' str_생년월일 '.' str_근무처
텍스트 대체
str_PDF결과 에서 '' 텍스트를 "(으)로 대체하고 결과를 str_PDF결과 에 저장
새 PDF 파일로 PDF 파일 추출
CurrentItem PDF 문서에서 str_바탕화면 \RPA_MSPAD\결과' str_PDF결과 '.pdf' PDF 파일로 1 페이지 추출 및 추출된 PDF를 v2페이지를삭제한PDF 에 저장
End 끝
폴더의 파일 가져오기
str_바탕화면 \RPA_MSPAD\결과' 폴더에서 '*' 필터와 일치하는 파일을 검색하고 Files 에 저장
PDF 파일 병합
Files PDF 파일을 str_바탕화면 \RPA_MSPAD\결과\병합.pdf(으)로 병합 및 병합된 PDF를 MergedPDF 에 저장
ZIP 파일
파일/폴더 str_바탕화면 \RPA_MSPAD\결과'를 str_바탕화면 \RPA_MSPAD\PDF데이터추출결과.zip'에 압축하고 ZipFile 에 저장