

캡스톤디자인 중간보고서

제 목	국문		논문 RPA 웹 서비스	
	영문		RPA Web Service for Paper	
진 행 상 황	중요마일스톤	<p><주요 기능 : 논문 RPA ></p> <p>장비 요구 사항</p> <ul style="list-style-type: none">- AWS의 EC2, S3을 활용한 Web Site 및 DB 구현- Vercel 서비스를 활용한 RPA모델 배포 <p>기능 요구 사항</p> <ul style="list-style-type: none">- Python 라이브러리를 활용한 문서 내 표 추출 기능- 오픈소스 모델을 통한 요약, 번역 등 문서 처리 기능 <p>성능 요구 사항</p> <ul style="list-style-type: none">- 기존 모델의 문서 처리 알고리즘 개선을 통한 사용자 요구 처리 시간 향상 <p>인터페이스 요구사항</p> <ul style="list-style-type: none">- 파일 첨부 및 정의한 변환 서비스 선택(사용자 인터페이스로 구현)- 사용자의 POST 메시지 내용을 토대로 웹/모델 통신에서 처리(서버간 인터페이스)- 사용자가 업로드한 원본 및 변환 결과물 저장(웹/DB 인터페이스)		
	진행상황	<ul style="list-style-type: none">- 웹 사이트 내 문서 업로드 UI 구현 완료- 업로드 원본 및 결과물을 저장하기 위한 AWS-S3(원격 DB) 연결 완료- 업로드한 문서 내 표 요소 추출 기능 구현 완료		
산출물	요구사항 정의서(별첨 1), 중간보고서(별첨 2)			
팀 구성원	학년	학 번	이 름	연락처(전화번호/이메일)
	4	20197128	이영호	010-8313-3747 / 20197128@edu.hanbat.ac.kr
	4	20181597	이태운	010-8486-5684 / 20181597@edu.hanbat.ac.kr
	4	20217134	박민이	010-9672-1996 / 20217134@edu.hanbat.ac.kr
컴퓨터공학과의 프로젝트 관리규정에 따라 다음과 같이 요구사항 정의서와 중간보고서를 제출합니다				
2024 년 5 월 2 일				
책임자 : 이영호 (인)				
지도교수 : 이현빈 (인)				

[별첨1]

프로젝트명 : 논문 RPA 웹 서비스

소프트웨어 요구사항 정의서

Version 1.0

개발 팀원 명(팀리더):이영호

이태윤

박민이

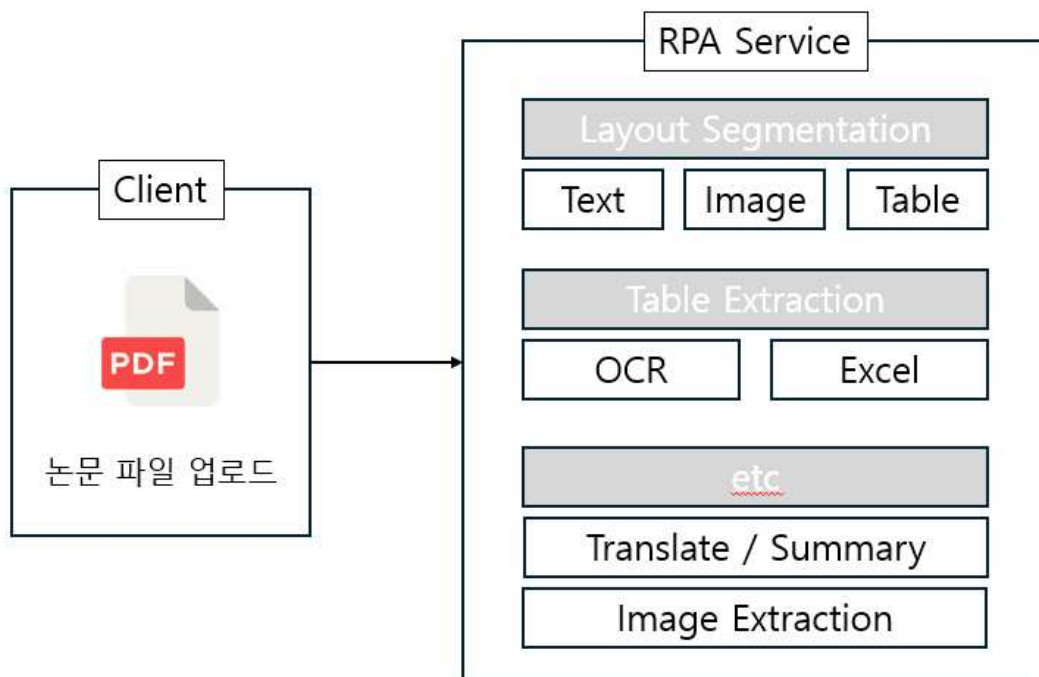
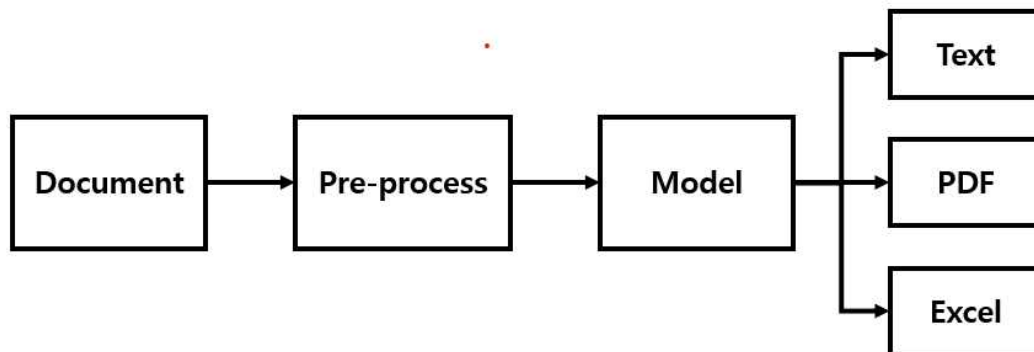
대표 연락처:010-8313-3747

e-mail: 20197128@edu.hanbat.ac.kr

목차

1. 개요
2. 시스템 장비 구성요구사항
3. 기능 요구사항
4. 성능 요구사항
5. 인터페이스 요구사항
6. 데이터 요구사항
7. 테스트 요구사항
8. 보안 요구사항
9. 품질 요구사항
10. 제약 사항
11. 프로젝트 관리 요구사항

1. 시스템 개요



2. 시스템 장비 구성요구사항

요구사항 고유번호		ECR-001		
요구사항 명칭		웹 서버		
요구사항 분류		시스템 장비구성 요구사항	응락 수준	필수
요구사항 상세설명	정의	<ul style="list-style-type: none"> - 웹 서버 		
	세부 내용	<ul style="list-style-type: none"> - 장비 품목: AWS EC2 - 장비 수량: 1식 - 장비 기능: 서버 - 장비 성능 및 특징: <ul style="list-style-type: none"> 1) 웹 서버 기능 및 서버간 데이터 처리 기능 2) DB를 통한 데이터 저장 		

요구사항 고유번호		ECR-002		
요구사항 명칭		데이터 서버		
요구사항 분류		시스템 장비구성 요구사항	응락 수준	필수
요구사항 상세설명	정의	<ul style="list-style-type: none"> - 파일 서버 		
	세부 내용	<ul style="list-style-type: none"> - 장비 품목: AWS S3 - 장비 수량: 1식 - 장비 기능: 서버 - 장비 성능 및 특징: 파일 업로드 및 다운로드 기능 		

요구사항 고유번호		ECR-003		
요구사항 명칭		모델 서버		
요구사항 분류		시스템 장비구성 요구사항	응락 수준	필수
요구사항 상세설명	정의	<ul style="list-style-type: none"> - 모델 서버 		
	세부 내용	<ul style="list-style-type: none"> - 장비 품목: 미정 - 장비 수량: 1식 - 장비 기능: 서버 - 장비 성능 및 특징: <ul style="list-style-type: none"> 1) 문서 파일 전처리 2) 모델 및 알고리즘을 통한 기능 구현 		

3. 기능 요구사항

요구사항 고유번호		SFR-001		
요구사항 명칭		문서 원본 및 변환 데이터 저장		
요구사항 분류		기능	응락수준	필수
요구사항 상세설명	정의	<ul style="list-style-type: none"> - 웹 서비스 - 사용자 업로드 - AWS-S3 - 사용자가 웹 서비스에 업로드한 원본 및 변환 문서를 AWS-S3를 통해 원격으로 저장하는 기능 		
	세부 내용	<ul style="list-style-type: none"> - 원격 저장소의 폴더 구조를 원본 및 변환된 문서 기준으로 분류하여 관리하는 기능 - AWS-S3의 정책 설정을 통해 업로드한 당사자만 해당 문서에 접근할 수 있는 보안 기능 - 업로드한 파일을 사용자가 원할 때 받아올 수 있는 불러오기 기능 		

요구사항 고유번호		SFR-002		
요구사항 명칭		문서 RPA 모델 개발		
요구사항 분류		기능	응락수준	필수
요구사항 상세설명	정의	논문 구성 요소 기반 변환 모델 개선		
	세부 내용	<ul style="list-style-type: none"> - DNN 기반의 Layout 분석 모델 개선 - 기존의 표의 구조상 발생하는 인식 문제(병합된 셀 등) 정확도 개선 - 문서 Layout 분석 속도를 개선해 사용자에게 더 나은 문서 처리 속도 제공 		

4. 성능 요구사항

요구사항 고유번호		PER-001		
요구사항 명칭		동시 접속 사용자 수		
요구사항 분류		성능 요구사항	응락 수준	필수
요구사항 상세설명	정의	웹 서비스 페이지, 문서 처리 모델 서버		
	세부 내용	<ul style="list-style-type: none"> - 동시 사용자 500명을 지원해야 하고 성능이 저하되지 않아야 함. - 처리된 파일이 다운로드 가능하게 된 후 30분 내의 시간 동안 Active 사용자로 간주함. - 문서 처리 모델 서버의 경우 시스템 부하가 80% 이상인 경우 동시 처리를 제한하고 우선순위를 부여하여 시스템이 마비되지 않도록 함. 		

요구사항 고유번호		PER-002		
요구사항 명칭		웹 페이지 디스플레이 시간		
요구사항 분류		성능 요구사항	응락 수준	필수
요구사항 상세설명	정의	웹 서비스 페이지		
	세부 내용	<ul style="list-style-type: none"> - 페이지 내의 요청에 따른 결과가 조회되는 것에 대한 응답 시간 - 시스템이 만드는 웹 페이지의 경우, 사용자가 요청한 시각으로부터 2초 내에 완전히 디스플레이 되어야 함 - 예외사항: 이 요구사항은 문서 업로드에 따른 처리 결과 파일의 다운로드까지의 과정에는 해당하지 않음 		

요구사항 고유번호		PER-003		
요구사항 명칭		처리 속도 및 시간		
요구사항 분류		성능 요구사항	응락 수준	필수
요구사항 상세설명	정의	문서 RPA 서비스		
	세부 내용	<ul style="list-style-type: none"> - 문서 처리의 경우 파일의 용량, 페이지 수에 따라 시스템 과부하가 발생할 확률이 높음. - 예상 시간, 처리 시간을 디스플레이 함 - 문서 처리 후 다운로드 가능할 때 알림을 통해 사용자의 대기를 감소토록 함. 		

요구사항 고유번호		PER-004		
요구사항 명칭		오류 응답 시간		
요구사항 분류		성능 요구사항	응락 수준	필수
요구사항 상세설명	정의	웹 서비스 페이지, 파일 서버, 모델 서버		
	세부 내용	<p>- 아래의 예상되는 오류에 대하여 2초 이내에 오류메세지를 전달해야 함.</p> <ol style="list-style-type: none"> 1) 잘못된 포맷의 파일 입력 2) 파일 서버 오류 3) 모델 서버 오류 		

5. 인터페이스 요구사항

요구사항 고유번호		SIR-001		
요구사항 명칭		드롭박스		
요구사항 분류		사용자 인터페이스	응락 수준	필수
요구사항 상세설명	정의	- 파일 업로드 드롭박스 구현		
	세부 내용	<ul style="list-style-type: none"> - 사용자가 서비스를 받고 싶은 파일을 업로드할 수 있도록 웹 페이지 홈 화면에 드롭박스 UI를 구성함. - 파일 확장자를 PDF로 제한을 둠. 해당 확장자가 아니면 사용자에게 에러 메시지를 제공함. 		

요구사항 고유번호		SIR-002		
요구사항 명칭		기능 템플릿		
요구사항 분류		사용자 인터페이스	응락 수준	필수
요구사항 상세설명	정의	- 사용자가 원하는 서비스가 담긴 템플릿 제공		
	세부 내용	<ul style="list-style-type: none"> - 사용자가 업로드한 파일을 토대로 해당 기능을 사용할 수 있는 템플릿 리스트를 홈 화면에 제공함. - 템플릿 페이지 안에는 산출물을 다운로드 받을 수 있는 버튼 UI를 구성함. 		

요구사항 고유번호		SIR-003		
요구사항 명칭		태스크 모달		
요구사항 분류		사용자 인터페이스	응락 수준	필수
요구사항 상세설명	정의	- 사용자에게 진행상황을 알리는 모달을 하단에 띄움		
	세부 내용	<ul style="list-style-type: none"> - 사용자가 선택한 템플릿에 따른 태스크의 진행상황을 알리는 모달을 띄워 작업시간이 얼마나 남았는지를 표시함. - 작업 완료 후, 사용자에게 알려 산출물을 다운로드 받게 함. 		

6. 데이터 요구사항

요구사항 고유번호	DAR-001		
요구사항 명칭	논문 자료 수집		
요구사항 분류	데이터	응락 수준	필수
요구사항 세부 내용	<ul style="list-style-type: none"> - 표 추출 및 변환, 텍스트 번역 및 요약 서비스의 테스트를 수행하기 위해 논문 PDF 문서를 사용함. - 논문 PDF 문서 안에는 그림, 표, 제목, 텍스트가 포함되어 있음. 		

요구사항 고유번호	DAR-002		
요구사항 명칭	문서 관리		
요구사항 분류	데이터	응락 수준	필수
요구사항 세부 내용	<ul style="list-style-type: none"> - 사용자가 업로드 한 문서 파일들을 AWS E2C에 저장 후 FASTAPI를 통해 기능에 맞게 변환함. - 문서 파일들은 외부의 다른 클라이언트나 서버에서 접근 불가함. 		

7. 테스트 요구사항

요구사항 고유번호	TER-001		
요구사항 명칭	테스트 방안		
요구사항 분류	테스트	응락 수준	필수
요구사항 세부 내용	<p>- 시스템 테스트에서 베이스라인이 되는 논문 자료를 사용하여 실제 환경에서 테스트를 수행함으로써 성능을 평가함.</p> <p>각 기능 테스트</p> <p>1) 문서에서 표를 올바르게 인식하고 추출하는지, 추출된 표가 엑셀 파일 형식으로 알맞게 변환하는지 확인함.</p> <p>2) 논문이 영어로 표기되어있는 경우, 번역했을때 본래의 뜻에서 벗어나지 않게 한글로 번역이 되었는지 확인함.</p> <p>3) 각 기능들이 통합되어 원활하게 작동되는지 확인함.</p>		

8. 보안 요구사항

요구사항 고유번호		SER-001		
요구사항 명칭		응용 및 DB보안		
요구사항 분류		보안	응락수준	필수
요구사항 상세설명	세부 내용	<ul style="list-style-type: none"> - DB 정책 설정을 통한 업로드 당사자 기준으로 접근 제한 - 저장된 파일의 이름 암호화 - DB 관리자 계정은 MFA인증을 통해 지정하지 않은 타인이 무단으로 도용하지 못하게 예방 		

9. 품질 요구사항

요구사항 고유번호		QUR-001		
요구사항 명칭		기능 통합 관리		
요구사항 분류		품질	응락 수준	필수
요구사항 상세설명	정의	- 품질 관리 (기술 관점)		
	세부 내용	<ul style="list-style-type: none"> - 각 템플릿에 해당하는 기능들이 통합되어 오류가 발생하지 않는지 테스트 하며, 오류 발생률을 최대 5%로 제한을 둠. - 사용자가 파일을 업로드하고 원하는 형식의 문서로 다운로드 받을 때까지의 서비스 처리 시간을 10분 이내로 둠. - 기능 서비스를 진행할 때, 오류 발생시 로깅되고, 시스템 관리자 및 사용자에게 알려 신속하게 대응할 수 있게 함. 또한, 오류의 원인을 분석하고 향후 발생을 방지하기 위한 조치를 취할 수 있도록 오류에 대한 상세한 정보를 제공함. 		

요구사항 고유번호		QUR-002		
요구사항 명칭		산출물 관리		
요구사항 분류		품질	응락 수준	필수
요구사항 상세설명	정의	- 품질 관리		
	세부 내용	<ul style="list-style-type: none"> - 시스템이 생성하는 모든 산출물(pdf 표 이미지, 번역된 텍스트, 본문 요약)은 오류나 왜곡이 없어야 하며, 사용자가 요구하는 목적을 충족시켜야 함. - 각 산출물은 수명 주기를 가지며, 필요한 경우 서비스 내에서만 운용이 가능하게 만들고, 사용자가 서비스를 종료했을 경우에는 삭제하여 산출물의 축적을 방지함. 		

10. 제약 사항

요구사항 고유번호		COR-001		
요구사항 명칭		시스템 개발과 설계 및 구현 제약사항		
요구사항 분류		제약사항	응락 수준	필수
요구사항 상세설명	세부 내용	<ul style="list-style-type: none"> - 현재 보유하여 활용 가능한 H/W, S/W 지식을 최대한 활용하여 진행 - 프로젝트에 필요한 라이브러리, 프레임워크 버전은 LTS 버전으로 통일하여 진행 		

11. 프로젝트 관리 요구사항

요구사항 고유번호		PMR-001		
요구사항 명칭		품질 관리		
요구사항 분류		프로젝트 관리	응락 수준	필수
요구사항 상세설명	세부 내용	<ul style="list-style-type: none"> - 프로젝트 과정에서 생성되는 작업 단위별 산출물에 대하여 작성 및 관리 - 개발 및 보고서 제출 		

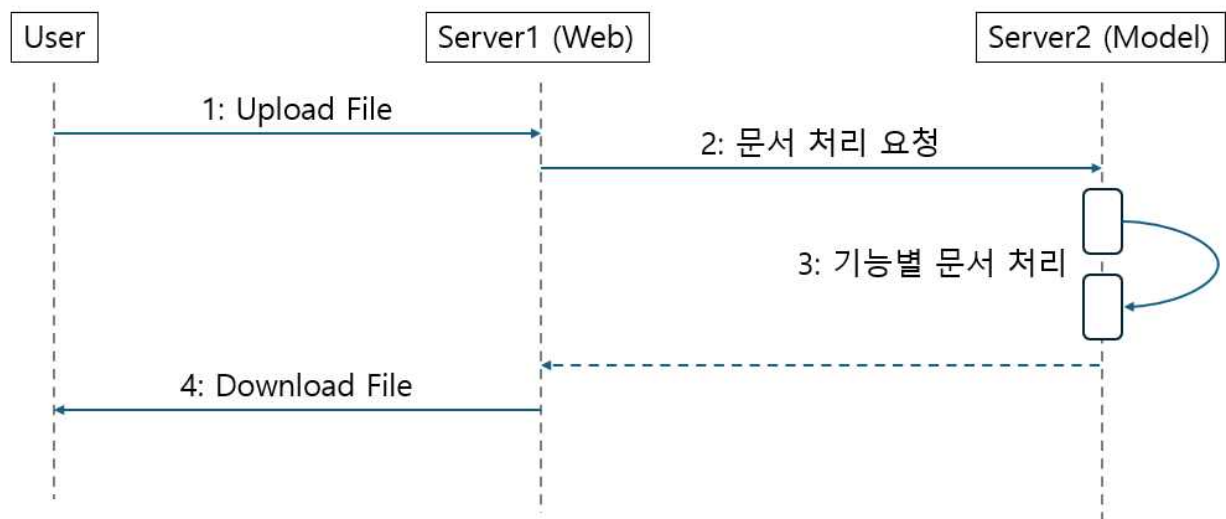
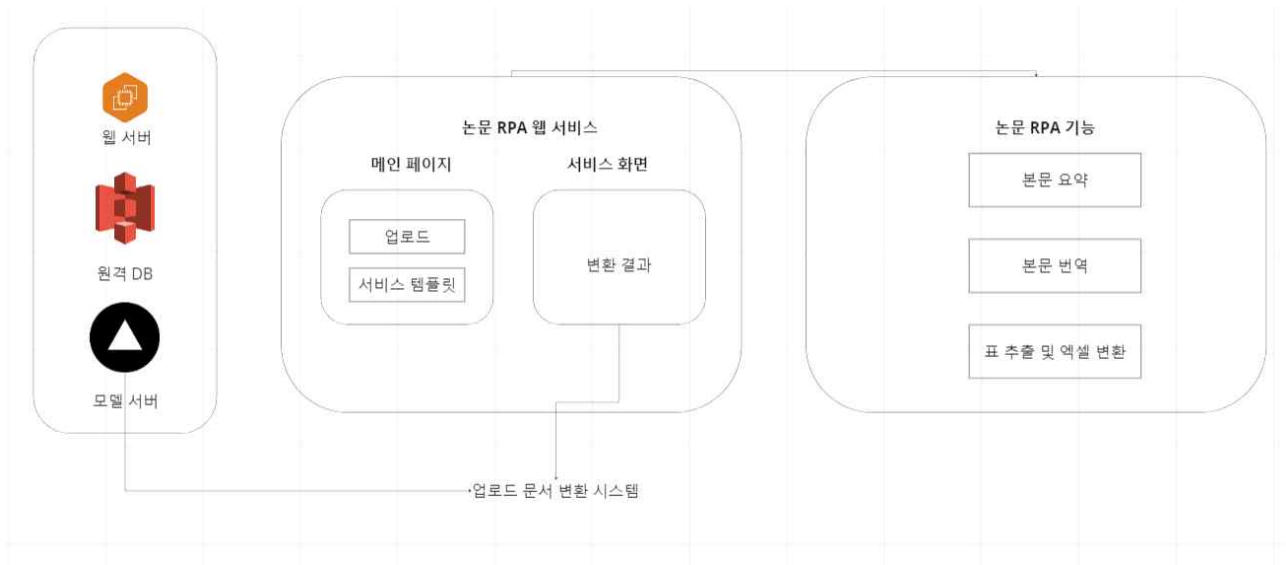
요구사항 고유번호		PMR-002		
요구사항 명칭		프로젝트 관리		
요구사항 분류		프로젝트 관리	응락 수준	필수
요구사항 상세설명	세부 내용	<ul style="list-style-type: none"> - 프로젝트 일정 계획에 따라 주1회 오프라인으로 진행 과정 공유 		

[별첨2]

중간보고서

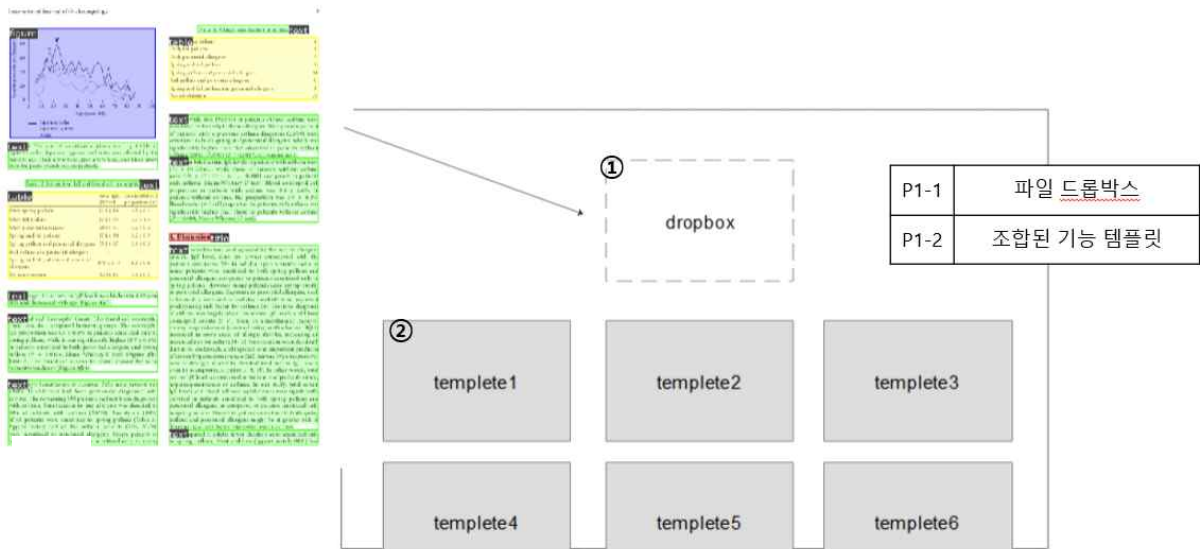
1. 요구사항 정의서에 명시된 기능에 대하여 현재까지 분석, 설계, 구현(소스코드 작성) 및 테스트한 내용을 기술하시오.

<시스템 구성도>



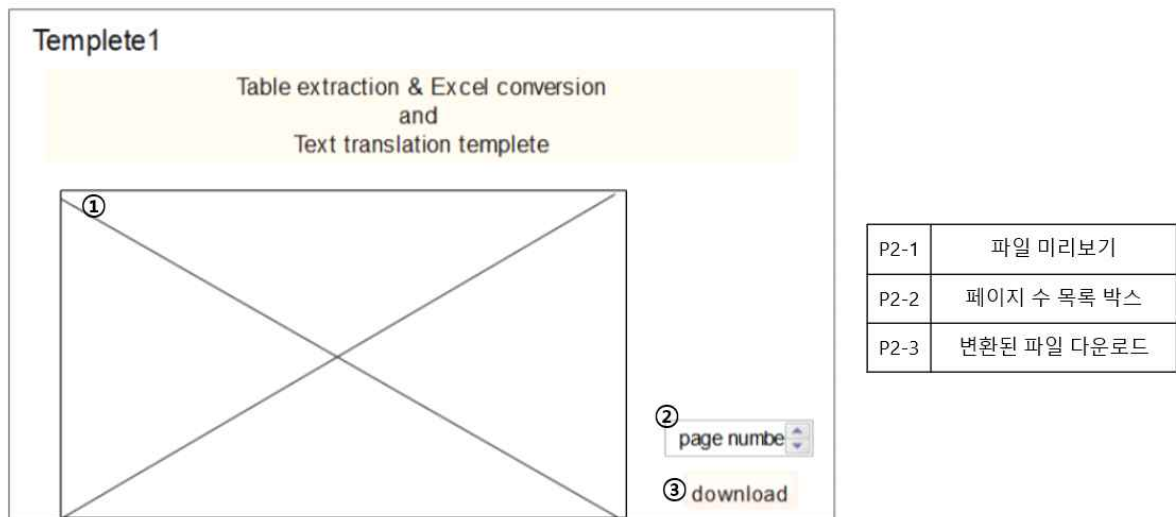
<Mock-Up>

P1



- 웹 페이지 메인 화면 부분으로 드롭박스과 표 추출, 엑셀 변환, 텍스트 요약, 번역 등의 기능의 조합으로 생성된 템플릿 블록들을 구성함.

P2



- 사용자가 템플릿을 선택했을 때 나오는 페이지로 이 템플릿같은 경우에는 표 추출, 엑셀 변환과 텍스트 번역의 기능이 담겨져 있음. 사용자는 다운로드 받을 파일을 미리 보기로 확인 가능하며, 페이지 수를 목록 박스로 원하는 문서 페이지를 선택해 다운로드 받을 수 있음.

2. 프로젝트 수행을 위해 적용된 추진전략, 수행 방법의 결과를 작성하고, 만일 적용과정에서 문제점이 도출되었다면 그 문제를 분석하고 해결방안을 기술하시오.

1) 초기 프로젝트 계획이 미흡하다는 평가를 받음.

- 초기 프로젝트 계획인 ‘표 추출하여 엑셀 변환’을 기능으로 한 RPA 웹 서비스로 계획을 확장하고 프로젝트의 대상을 논문을 활용하려는 학생으로 한정하여 서비스 개발 목적과 방향성을 수립

2) 개발을 진행하면서 소스코드의 결합도로 인해 업데이트 및 유지보수에 어려움을 겪음

- 코드 작성 전에 개발할 기능과 작업 분할을 미리 회의를 통해 구분하여 작업 간의 충돌과 반복을 감소시킬 수 있었음.
- git branch를 통한 기능 개발 시 branch의 길이를 길게 가져가지 않고 반복적인 merge를 통해 업데이트 및 유지보수를 즉각적으로 진행할 수 있었음
- Notion 페이지를 새로 생성해서 구현해야할 기능들과 구현한 기능을 정리하여 개발하고, 또한 각자 맡은 파트에서 문제가 발생할 경우 이슈 관리 페이지를 통해 바로바로 소통을 계속 이어나가면서 개발을 진행할 예정임.

3) 학기가 진행되면서 캡스톤 프로젝트 일정계획이 지연되었음.

- 실제 구현되는 서비스의 개발이 중요하다고 생각하여 세부 사항과 기능들에 우선순위를 부여하여 서비스 개발의 속도를 높이하고자 함.
- 또한, 온라인을 통한 작업 공유에 한계가 있다고 생각하며 오프라인 환경에서 진행 과정을 공유하고 함께 개발하는 시간을 가져 빠른 개발을 도모하고자 함.