

# 결혼 관련 계약서 자동 정리 서비스 실무 기획서

## 1. 프로젝트 개요 및 목표

### 1) 목적

결혼 준비 과정에서 발생하는 계약서, 견적서, 확인서 등 다양한 서류들을 자동으로 정리하고, 스캔된 이미지나 PDF 형태의 문서를 OCR 기반으로 텍스트화하여 항목별로 분류 및 요약하는 서비스를 구축해서 파일을 열어보지 않아도 주요 계약 내용(금액, 업체명, 일정 등)을 한눈에 확인할 수 있으며, 서류 관리의 번거로움을 줄이고 계약 조건 비교나 일정 관리까지 자동화 가능

### 2) 기대효과

- 서류 관리 시간 80% 이상 단축
- 계약 조건 및 일정 관리 자동화
- 문서 누락 및 중복 관리 개선
- 결혼 준비 과정의 스트레스 완화 및 효율성 향상

## 2. 요구사항 정의 및 분석

### 1) 대상 문서 특성

- 인쇄된 계약서, 견적서, 영수증 등 스캔된 이미지(PNG/JPG) 및 PDF 파일
- 결혼 관련 서류 중심 (예: 음식점 계약서, 드레스샵 견적서, 스튜디오 촬영 계약서, 메이크업샵 예약확인서 등)
- 한글 위주, 일부 영문/숫자 혼용 가능
- 양식 불균일 (업체마다 문서 포맷 다름)
- 핵심 정보가 표 또는 문장 형태로 존재 (날짜, 금액, 업체명, 조건 등)

### 2) 중요 고려사항

- 스캔 상태나 인쇄 품질이 다른 다양한 서류를 안정적으로 인식
- 문서 제목 또는 내용 기반으로 “음식장”, “드레스”, “촬영”, “메이크업” 등 카테고리 분류
- 금액, 업체명, 계약일, 예약일 등 주요 정보 자동 인식
- OCR 결과와 원본 파일을 함께 보관 및 검색 가능
- 이름, 연락처 등 민감정보 자동 마스킹 기능 필요

## 3. 기술 구성 및 아키텍처

### 1) 이미지 전처리 및 OCR 처리

- OpenCV 기반의 grayscale 변화, adaptive thresholding
- 기울기 및 왜곡 보정 (Hough Transform / contour alignment)
- 해상도 보정 (DPI 미달 시 보간 업스케일링)
- OCR 엔진 처리
  - Tesseract (LSTM 기반 한글 + 영문)
  - HuggingFace TrOCR 한글 모델 파인튜닝 적용
  - Confidence Threshold 0.85 미만 문장은 후처리 재검증

### 2) 텍스트 분석 및 핵심 필드 추출

- Python regex, dateparser, spacy (Ko 모델)
- 추출된 데이터는 DB에 key-value 형태로 저장

### 3) 문서 자동 분류

- 1차 키워드 매칭
- 2차 문서 임베딩 분류(Sentence-BERT 기반 임베딩 후 cosine similarity 분류)

#### 4. 보안 및 개인정보 보호

- 모든 업로드 파일 AES-256 암호화 저장
- HTTPS 기반 TLS 암호화 전송
- 이름, 연락처, 계좌번호 등 자동 마스킹 처리
- 개인 로그인 기반 접근 제한, JWT 토큰 인증

#### 5. 프로젝트 추진 계획 및 조직

##### 1) 개발 구성

- 프로젝트 관리자 : 서비스 기획, UI 설계, 전체 일정 관리
- 백엔드 개발 : 파일 업로드, OCR 처리, 데이터 저장/조회 API 구현
- 프론트엔드 개발 : 문서 업로드, 요약 결과, 검색/필터 UI 개발
- ML 엔지니어 : OCR 모델 및 텍스트 추출 로직 설계
- 보안 관리 : 데이터 암호화, 접근 제어 로직 구축

##### 2) 개발 일정

- 요구사항 정리 및 UI 시안 제작 (1개월)
- OCR 엔진 적용 및 필드 추출 로직 구현 (1개월)
- 문서 분류 및 검색 기능 개발 (1개월)
- 테스트 및 배포, 사용자 피드백 반영 (1개월)

#### 6. 기대 효과 및 향후 확장성

##### 1) 기대 효과

- 서류 찾기 및 관리 시간 대폭 절감
- 파일 열람 없이도 핵심 정보 즉시 확인 가능
- 결혼 준비 과정의 정보 정리 효율 극대화
- OCR 및 NLP 기술의 실무 적용 경험 확보

##### 2) 향후 확장성

- 가계약/잔금 일정 자동 알림 기능
- 업체별 조건 비교 및 시각화
- 모바일 앱 버전 출시 (사진 촬영 후 자동 등록)
- 부부 공동계정 공유 기능
- 다른 영역(이사, 여행, 보험 등)으로 확장 가능

# 결혼 관련 계약서 자동 정리 서비스 실무 기획서(1주일)

## 1. 프로젝트 개요 및 목표

### 1) 목적

결혼 준비 과정에서 발생하는 다양한 계약서(PDF, 이미지 등)를 AI OCR 기술로 자동 인식하여 핵심 정보를 정리하는 문서 정리 도우미 서비스를 구현

### 2) 범위

- 스캔된 계약서 이미지/PDF 업로드
- Tesseract OCR로 텍스트 추출
- 계약 금액·업체명·날짜 등의 핵심 정보 자동 파싱
- 결과를 웹 페이지에서 확인

## 2. 요구사항 정의 및 분석

### 1) 입력 문서 특성

- 파일 형식: pdf, jpg, png
- 주요 내용: 결혼 관련 계약 정보(예식장, 드레스샵, 스튜디오 등)
- 한글 중심 문서, 일부 숫자·영문 혼용

### 2) 기능 요구사항

- 사용자가 계약서(PDF/이미지)를 업로드
- 업로드된 파일에서 텍스트 인식
- 금액, 날짜, 업체명 등 정규식 기반 추출
- 인식된 텍스트 및 요약 정보 웹 화면에 표시
- 최근 업로드 문서 리스트 조회 기능

## 3. 기술 구성 및 아키텍처

### 1) 이미지 전처리 및 OCR 처리

- 전처리(OpenCV 사용)
  - Grayscale 변환으로 글자 대비 향상
  - Adaptive Thresholding으로 조명 불균일 보정
  - 간단한 기울기 보정 (Hough Line Transform 사용)
  - PDF 입력 시 pdf2image 라이브러리로 페이지별 이미지 변환
- OCR 엔진 처리
  - Tesseract (LSTM 기반 한글 + 영문)
  - Confidence score(인식 신뢰도) 출력

### 2) 텍스트 분석 및 핵심 필드 추출

- Python 기반 단순 정규식 매칭 (Regex)

### 3) 문서 자동 분류

- 키워드 매칭

## 4. 보안 및 개인정보 보호

- 업로드 파일은 임시 폴더(/tmp)에 저장 후 삭제
- HTTPS 기반 TLS 암호화 전송
- 이름, 연락처, 계좌번호 등 자동 마스킹 처리

## 5. 프로젝트 추진 계획 및 조직

### 1) 개발 구성

- 프로젝트 관리자 : 서비스 기획, 기술 스택 선정, 일정 관리, 전체 코드 구현
- 백엔드 개발 : Flask 기반 API, OCR 처리 및 텍스트 분석 로직 구현
- 프론트엔드 개발 : 업로드·결과 표시용 간단한 웹 인터페이스 구축
- AI/OCR 엔지니어 : Tesseract OCR 적용 및 핵심 필드 추출 알고리즘 설계

### 2) 개발 일정

- 기획 및 환경 세팅 (1~2일)
- 핵심 기능(텍스트 인식 및 정보 추출 기능 동작) 구현 (3~5일)
- 결과 페이지 및 테스트 (6~7일)

## 6. 기대 효과 및 향후 확장성

### 1) 기술적 효과

- 실제 계약서 스캔본에서 한글 OCR 인식을 검증
- 정규식 기반 핵심 필드 추출이 실제 서류에서 얼마나 유용한지 평가
- Flask를 이용한 파일 업로드-처리-출력 파이프라인 구성 경험
- 향후 TrOCR, Vision API 등으로 성능 개선이 가능한 구조 확보

### 2) 개인적 성장 효과

- 프로젝트 기획~개발 전 과정 경험
- OCR/NLP 전처리 기술에 대한 실무 감각 습득
- 실제 생활 불편함을 해결하는 기술 서비스의 초기 모델 검증

### 3) 향후 기대 효과

- 결혼 서류뿐만 아니라 계약서, 견적서, 세금계산서 등 문서 자동 관리 시스템으로 확장 가능
- 향후 AI 기반 문서 검색·요약 플랫폼으로 발전 가능
- 개인 데이터 관리 서비스(Productivity Tech) 분야로 연결될 가능성