**아이디어 기획서**

| **팀명** | 키득키득 |
| --- | --- |
| **팀원 및 역할** | ▪팀장 : 홍길동 (프로젝트 총괄, 일정계획, Back End , Model)  ▪팀원 : 김가윤 (Front End, 서류작업) 장건 (Back End, Model)  김청래 (Front End, 서류작업) 조호현 (Back End, Model)  하영진 (Front End, 서류작업) |
| **아이디어 주제** | 많은 서류의 내용을 정리하는 작장인들을 위한 객체 인식과 OCR을  통해 제공받은 키워드로 서류의 중요 자료를 정리하는 서비스 |
| **제안 배경 및**  **필요성** | ▪ 시장현황  전년도보다 수출 수입의 규모가 커짐에 따라 거래량이 증가하면서 그에 따른 거래문서 수량이 증가함.  ▪ 현대 사업의 대부분의 업무 지행을 디지털 기반에서 수행하지만, 외부와의 협업관계에 필요한 계약 및 증빙 문서는 hard-copy 종이 문서를 사용하는 형태가 많다.  ▪ 서류 정리의 시간을 줄이고 문서의 내용의 핵심을 정리해 주어 빠른 업무 처리가 가능하다. |
| **유사 제품 현황**  **및 비교** | ▪ (카카오 테크그라운드 지문분석) |
| **제안 내용** | ▪ 개발 목표   * 서류의 내용을 사람이 인지하여 key-in 하지 않고 문서 내용을 MRC를 통해 자동 key-in 할 수 있도록 진행   ▪ 개발 내용    문서를 업로드 시 해당 데이터를 OCR, Deep Learning을 통해 인식하여 문서화 시키는 웹 기반 프로그램 제작 |
| **수행 방법** | ▪데이터 확보방안  제안 업체에서 여러 종류(.XML, .PDF 등)의 데이터를 제공 예정  ▪추진 전략( 일정, 수행 방법 등 )  PDF파일을 입력 시 테서렉트를 통해 OCR분석하여 키워드 기반 요약 내용 산출, 웹 기반 프로그램을 통해 가동 |
| **기대효과 및 활용방안** | ▪기대 효과  업무 단순화 및 사용자의 편리성 증가로 작업시간 단축  ▪활용 방안  방대한 양의 서류를 키워드 기반으로 단시간에 정리 가능  서류를 키워드 별로 관리가 가능 |

**(1) 제안 배경 - 외부 환경 분석 (PEST / STEEP)**

| **정책적 배경** | ▪정부에서 사회적 관심이 높은 기후변화, 미세먼지, 폐기물, 화학물질 등 환경 이슈에 주도적으로 대응해 나가고 있습니다. 기업시민헌장, 환경경영방침, 관련 사규 및 내부지침에 의거하여 환경 가이드라인을 수립하고 사업/생산 과정에서 환경에 미치는 부정적 영향을 최소화하기 위하여 다음의 활동을 추진하고 있다. |
| --- | --- |
| **경제적 배경** | ▪ 2021년 조사에 따르면 2020년 국내 전자문서 공급시장 규모, 전년  대비 약 4천억 원 증가<산업일보 중 https://www.kidd.co.kr/news/225447> |
| **사회적 배경** | ▪ 코로나 19로 시작된 비대면화 현상으로 인해 주요 전자문서사업의 수요처 1위(49.8%)에 해당하는 분야는 ‘정보통신업 (소프트웨어 · 영상 · 방송 · 통신 · 정보 · 출판 )’인 것으로 나타났다. 제조업(14.4%), 전문, 과학 및 기술서비스업(10.8%), 도매 및 소매업(6.1%) 등은 뒤를 이었고 계속해서 전자 문서가 쏟아지고 있다. |
| **트렌드 배경** | ▪ Zero Paper  롯데건설은 이번 MS Teams 도입을 통해 임직원, 파트너사 간의 실시간 소통 기반을 마련해 비대면, 페이퍼리스(종이 없는) 업무 환경으로의 전환을 목표로 한다.  < 기사 중 <https://www.fetv.co.kr/news/article.html?no=126312>>  ▪ MRC 트랜드  COVID-19 팬데믹을 계기로 Digital Workplace의 필요성이 대두되면서 회사에 산재되어 있는 다양한 양질의 정보들을 하나로 통합하고, 필요한 내용을 검색하여 적시에 인사이트를 얻는 일이 보다 중요해지고 있습니다. 이러한 요구에 따라 대용량 문서를 대상으로 사용자의 질문에 똑똑하게 대답하는 Question Answering(QA) 시스템에 관한 관심이 커지고 있습니다. 필요한 정보를 빠르고 정확하게 찾는 기계독해, 시맨틱 검색 등의 자연어 처리와 자연어 이해 기술의 필요성이 부각되고 있는 이유입니다.  <https://www.samsungsds.com/kr/insights/techtoolkit\_2021\_qa.html> |
| **기술적 배경** | ▪ BERT  Representations from Transformers는 Google에서 개발한 자연어 처리 사전 교육을 위한 변압기 기반 기계 학습 기술  ▪ MRC  인공지능(AI) 알고리즘이 스스로 문제를 분석하고 질문에 최적화된 답안을 찾아내는 기술 |
| **제도적 배경** | ▪전자문서 및 전자거래기본법상 전자문서의 법적효력 및 서면요건을 명확히하고, 전자화문서 보관시 종이문서 폐기 근거를 마련하며, 온라인 등기우편을 활성화하기 위한 기반을 마련 하였습니다(’20. 6월 「전자문서 및 전자거래 기본법」 개정, ’20.12.10. 시행). ▣ 기존 종이문서를 스캔하여 변환한 전자문서를 공인전자문서센터\*에 보관하는 경우 해당 종이문서를 보관하지 않아도 되어 종이문서와 전자문서의 이중보관 부담이 해소됩니다. ＊전자문서의 안전한 보관 및 증명 등의 업무를 수행하기 위해 과기정통부장관으로부터 지정받은 전문기관 |

**(2) 제안 배경 - 내부 환경 분석 (3C)**

| **제안자 능력** | ▪ 플라스크 기반 모델 구축  ▪ Node.js 기반 웹 서버 구축  ▪ Tesseract-OCR을 통한 텍스트 추출 |
| --- | --- |
| **경쟁제품/기술/특허 분석** | ▪ 카카오 테크그라운드 지문분석  →지문을 입력하고 질문을 입력하면 지문을 기반으로 질문에 대한 적절한 답변을 출력 |
| **고객 분석** | ▪ 계약서가 많은 직종의 회사(제조업, 은행업) |

**(3) 필요성**

| **필요성** | ▪ 요즘 대다수의 거래는 전자문서로 진행되는데 이미지 형식으로 이뤄진 파일은 원하는 내용 찾기가 매우 번거로움. |
| --- | --- |
| **차별성** | ▪ 키워드 기반 분석으로 내용을 출력해줌. |
| **현재까지**  **준비사항** | ▪ Front End -> electron  ▪ Back End -> 모델 MRC Bert  Nodejs , Flask |
| **기대효과** | ▪ 사용자의 편리성 증가로 작업시간 단축 |
| **활용방안** | ▪ 다국적 기업의 장문의 거래문서를 단시간에 분석 후 요약내용 또는 정리 |

**(4) 개발 내용**

| **개발 목표** | ▪ 서류의 내용을 사람이 인지하여 Key-in 하지 않고 문서 내용을  비전인식 및 디지털화하여 자동으로 Key-in되도록 함. |
| --- | --- |
| **개발 내용** | ▪ Front End :  Electron 을 통한 웹 기반 프로그램  ▪ Back End :   * Pytesseract 를 사용한 PDF 문서 내용 인식 * MRC Berk를 사용한 딥러닝 모델 제작 * 모델서버 : Flask * 웹 프로그램 서버 : Node js |

**(5) 달성 목표 및 달성 전략**

| **달성 목표** | ▪ PDF 문서에 대한 인식 및 키워드를 통한 내용 추출 웹 기반 프로그램 |
| --- | --- |
| **달성 전략** | ▪ Front End :   * Electron 을 활용한 웹 기반 프로그램 제작 * 프로젝트 서류   ▪ Back End :   * Pytesseract 를 사용한 PDF 문서 내용 인식 * MRC-Bert를 사용한 딥러닝 모델 제작 * 모델서버 : Flask * 웹 프로그램 서버 : Node js |

**(6) 개발 일정**

**<참조 1>**

**(7) 참여 인원**

| **이름** | **역할 및 능력** |
| --- | --- |
| **고경훈** | ▪ PM |
| **김가윤** | ▪ 프론트 엔드 팀 총괄 |
| **김청래** | ▪ 프론트 엔드 팀원 + 주요 문서 제작 |
| **장건** | ▪ 백엔드 팀 총괄 |
| **조호현** | ▪ 백엔드 팀원 + 알고리즘 제작 |
| **하영진** | ▪ 프론트 엔드 팀원 |

**(8) 시장 분석**

| **이름** | **역할 및 능력** |
| --- | --- |
| **국내시장 규모 및 현황** | ▪ 한국인터넷진흥원, "2020년도 전자문서 산업 실태조사"에 따르면 전자문서시장은 총 3가지로 분류한다. 전자문서 생성·획득·변환업에는 2조 1044억원, 전자문서관리업에는 5조 1889억원, 전자문서 교환업에는 2조 3918억원으로 총 9조 6851억원으로 형성되어 있음. |

**(9) SWOT 분석**

| ▪ 전자 문서에 내용을 키워드에 대한 내용 추출로 인한 업무 노동량 축소 | **S** | **W** | ▪ 상정하지 못한 언어가 기입되어있으면 오류가 발생할 수 있음.  ▪ PDF가 아닌 다른 형식 파일을 사용할 수 없음 |
| --- | --- | --- | --- |
| ▪코로나로 인한 재택근무 활성화로 전자문서 확대  ▪ ESG로 인한 시장 확대 가능성 있음. | **O** | **T** | ▪ 현재 카카오AI에서 여러가지 MRC 모델과 OCR 모델이 데모 버전이 있음 |

**(10) STP 전략**

| **고객 분류** | ▪ 회사 및 공공기관 |
| --- | --- |
| **목표 고객** | ▪ 계약서가 많은 직종의 회사(제조업, 은행업) |
| **시장에서 제품의 위치** | ▪ 현재 카카오엔터프라이즈는 기술의 체험을 목적을 있으나 업무 제휴을 통한 기술을 나누어지고 있지만 우리는 업체별 다른 딥러닝을 통한 배포가 가능 |

**(11) 4P 전략**

| **제품** | ▪ 문서를 인식한 후 키워드 기반의 내용만 추출하는 웹 기반 프로그램 |
| --- | --- |
| **가격** | ▪ 구독 서비스로 진행하며 (매출액 기준으로 그룹을 나눠 가격책정) |
| **판매 방법** | ▪ 홈페이지를 통한 온라인 판매 |
| **홍보 방법** | ▪ 중소기업이 참석할만한 행사에 참석하여 홍보를 진행, 및 고객사의 진입장벽을 낮추기위한 프로모션을 진행하여 홍보할 계획 |