**PPAP(Patent Process Accelerating Program) Ideation 제안서**

|  |  |  |  |
| --- | --- | --- | --- |
| **제안서 제출자 성명** | 이정연 | **연락처** | 010-3622-9397 |
| **팀원 전원 정보\***  **(개인으로 참가 신청 시**  **“해당 없음”으로 기재)** | **참여자는 아래 4개 역할 중 하나를 필히 표기**   |  |  |  | | --- | --- | --- | | **역할** | **이름** | **전자우편/연락처** | | **디자인/Front-end/Interface** | 변아영 | [ayoungbyun@gmail.com](mailto:ayoungbyun@gmail.com) 010-7137-7328 | | **개발/로직/네트워크** | 조원익 | [wicho@hi.snu.ac.kr](mailto:wicho@hi.snu.ac.kr) 010-5414-1484 | | **데이터/Database** | 조호수 | [lakecho@snu.ac.kr](mailto:lakecho@snu.ac.kr) 010-8821-9499 | | **데이터 과학/알고리즘** | 이정연 | [curieuxjy@kakao.com](mailto:curieuxjy@kakao.com) 010-3622-9397 | | | | |

**\*팀으로 참가하는 경우 아이디어 제안서는 팀 대표 1명만 제출하면 되며, 팀원의 정보를 위와 같이 함께 기재해 주시기 바랍니다.**

***How might we utilize advanced analytical techniques to give our clients back the gift of time?***

인공지능기술(머신 러닝, 데이터 과학, 자동화, API 등)을 활용하여 낭비되는 시간을 대폭 줄여줄 수 있는 아이디어를 간단하게 기술해 주시기 바랍니다.

|  |
| --- |
| 아이디어 PPAP(Patent Process Accelerating Program)현황, 기존 경쟁 서비스 분석  * 1. [카카오 AI, 특허청 특허 심사에 활용된다](https://biz.chosun.com/site/data/html_dir/2020/01/09/2020010901473.html) (조선일보 기사) : 번역 서비스를 기반으로 특정 키워드를 담은 유사한 문서를 비교분석, 챗봇 서비스   2. [Patentability Search](https://sagaciousresearch.com/patentability-search/) : NLS(Native Language Search) 서비스 제공   3. [The patent search firm Services](https://thepatentsearchfirm.com/services/) : 검색이 아닌 훈련된 인력에게 보고서 작성을 의뢰하는 방식   4. [Patent search software - Dennemeyer Octimine](https://www.dennemeyer.com/octimine-patent-search-software/) : 문장형 검색 지원 / 검색 키워드 관련 특허 건수를 시각화해서 보여주는 등 고급 검색 기능을 지원   5. 기존 경쟁 서비스를 비롯하여 WIPS나 Kipris 또한 선행 특허를 검색하는 데에 집중  기존 경쟁 서비스와의 차별점과 필요성  * 1. 선행 특허 조사가 아닌 특허 서류를 작성하는 단계에 관여하는 최초의 서비스임   2. 특허 서류 작성시 자주 쓰이는 표현 및 단어를 추천하여, 초안 작성 시 특허 서류에서 지켜야 하는 엄격한 세부 규칙에 대해 발명인에게 알려줄 수 있음   3. 변리사가 새로운 아이디어에 관련된 학계의 동향 및 전문용어를 자세히 파악하고 있기 힘들다는 면에서, 관련 논문을 바탕으로 전문 용어를 알려주는 서비스는 단순 특허 서류 검토 및 대조로는 얻을 수 없는 효용을 가짐        * 변리사와 발명인 사이의 배경지식의 차이로 인해 불필요한 커뮤니케이션으로 인한 특허 출원 지연이 발생 * 변리사는 대개 기계공학분야, 화학공학분야, 전기전자분야, 바이오생명과학분야 중 한 개의 전문지식을 갖추고 활동하여 담당 분야 외의 지식이 없으며, 신기술 및 융합 기술이 적용된 발명 아이디어를 파악하는 데 어려움    아이디어  1. **유사 특허/논문 검색 서비스** 2. 문장 또는 문단으로 요약문을 입력하면 관련도가 높은 순서대로 특허나 논문 전문을 결과로 제시 3. 기술적으로 유사도가 높은 부분을 특정하여 강조해주는 점이 기존 검색과 차별화 4. 새로운 고안과 대조하여 관련 논문 및 특허 문서를 효율적으로 파악하는 데 도움 5. **특허 서류 에디터 (표현 검토 서비스 / 전문용어 해설 및 문장 추천 서비스)** 6. 발명인이 특허 문서의 초안을 입력하면 관련 특허/논문에서 사용빈도가 낮은 표현, 또는 특허 문서를 작성하는 데 지켜야 하는 형식적, 내용적 규칙을 어긴 표현에 대하여 사용자의 의도에 맞는 대체 표현을 추천 7. 관련된 논문 또는 특허 문서의 단어와 발명인이 초안에 사용한 단어를 매칭하여 작성 중인 문서 내 전문용어에 대한 이해를 돕는 설명 및 그림, 연관된 서술어, 다른 전문용어를 제시 8. 변리사가 해당 발명에 대한 선행 연구를 수집하고 특허 신청 서류 작성에 활용하는 데에 필요한 시간을 대폭 단축시키고, 최신 동향을 반영한 특허 서류를 작성할 수 있음 9. 기존 맞춤법 검사기와 같이 초안 전문을 입력하여 검토하는 방식, 또는 실시간 에디터형(Grammarly 서비스와 유사한 형태)으로 구현    기대효과  * 유사 특허 및 논문 검색의 정확도를 높이고 검색 경험을 원활히 함 * 변리사와 발명인 간의 정보 간극을 줄이면 특허 출원 시간과 비용을 획기적으로 감축 * 신기술 및 융합 기술이 적용된 고안에서 변리사가 전문용어를 이해하는 데 걸리는 시간을 절약 * 변리사가 관련 분야의 동향을 참조하여 발명인의 의도를 정확히 반영 * 발명인이 기등록된 특허 서류의 규칙과 관행을 파악하여 서류를 작성하는 데 도움 * 특허 분쟁을 감소 시킬 수 있고[[1]](#footnote-0), 특허 사냥꾼의 활동을 미연에 방지[[2]](#footnote-1) * 기술 기업의 연구개발 및 법무 부서에서 활용하여 이와 같은 특허 출원 지연으로 인한 분쟁 예방 및 비용 절약 |
| 스케치(그림) |
| 기술유사 특허/논문 검색 서비스  * 1. **특허 및 논문의 요약문 비교**   발명인이 작성한 제안서의 요약(abstract)을 기존의 특허나 논문의 요약과 비교함. 요약문을 바탕으로 하는 이유는 발명제안서 초안들이 대체로 특허 문건의 형태로 완성되어 있지 않기 때문이며, 여기에 논문들을 함께 활용하는 것은 분야별 최신 특허들이 논문들과도 밀접한 연관이 있기 때문.   * 1. **다언어 확장가능성**   한국어 문서를 가정하지만, 최소 영-한 code-mixed의 형태도 가능하게 구현 예정. 최신 기술에 활용되는 주요 단어들은 외국어 표현을 포함하는 경우도 종종 있기 때문. 이는 다언어로의 확장도 고려할 수 있는 방법론임.   * 1. **의미와 어휘의 종합적 활용**   요약문의 비교를 위해 어휘적인(lexical) 측면과 의미적인(semantic) 측면을 모두 고려함. 전자의 경우 사전 기반의 동시 출현 행렬(co-occurrence matrix)로부터 두 문서가 겹치는 정도를 구할 수 있음. 후자의 경우 KoBERT (distributed by SKT)와 같은 사전 학습된 언어모델로부터 그 수치적 표현을 얻을 수 있음. 따라서, 두 결과를 이용하여, 코싸인 유사도(spatial cosine similarity)나 최종 레이어 벡터의 코싸인 거리(cosine distance)를 계산한다면, 확실하게 함께 존재하는 용어처럼 관찰 가능한(observable) 비교부터 잠재적인(latent) 의미의 파악까지 점진적으로 수행할 수 있을 것으로 기대됨. 한국어 토크나이저 및 단어 사전은 상기 언어모델과 동일하게 구현 예정.   * 1. **문단 단위의 요약문 비교**   요약문의 각 문장보다는 문단 단위로 비교 예정이며, 이는 특정 품사만 골라서 분석에 사용하는 것은 특허 출원 서류의 독특한 형식이 난점으로 작용하기 때문임. 요약문도 일부보다는 전체를 비교에 활용할 예정이며, 주요 내용(proposal)과 부차적인 내용(background)의 대비는 주의(attention)에 기반한 알고리즘을 통해 보완할 예정.   * 1. **유사 특허의 반환 및 그 효과**   기술적 관련도/어휘적인 유사도를 비교 후, 사용자가 원하는 분야를 지정하여 top 10의 결과를 반환. 이로써, 함정 특허의 발견, 선행논문 검색, 유사 특허의 컨벤션 파악 등이 가능. 특허 서류 에디터  * 1. **기존 특허 및 논문의 표현 확인**   변리사(특허 작성자)가 발명인의 제안서에서 합의되지 않은, 또는 접하지 못했던 표현을 마주했을 때, 해당 표현(단어, 어구 및 어절 포함)이 기존의 유사 문건들에서 사용된 맥락을 보여줌   * 1. **표현을 이용한 직접 검색**   해당 표현이 직접 등장하는 부분을 검색하여 문맥을 파악하는 부분(named entity recognition, NER)으로, 일종의 Search by word로 볼 수 있음. 문맥(context) 및 연어(collocation)을 찾아주며, 편집 거리(edit distance)를 활용한 검색으로 구현   * 1. **문맥 파악**   의미/어휘적으로 유사할 가능성이 큰 부분에 분석을 집중하는 Search by context로써, 주변 단어, 어구/절, 해당 문장/문단 등으로 문맥의 범위를 지정할 수 있도록 하여, 후보 문건들로부터 비슷한 맥락에서 사용되는 표현들을 찾아 보여줌   * 1. **문맥 검색 및 그 효과**   발명인과 변리사 간 소통의 오해를 줄이고, 자체적인 오류 수정이 가능하게 하며, 문서 작성 시간을 절감할 수 있음. 또한, 분야를 넘나드는 신기술이 많이 등장하는 요즘, 서로 다른 분야 전문가들의 소통을 증진해 줄 수 있음. 데이터  * 1. **학습 데이터**      1. [AI Hub](http://www.aihub.or.kr/aidata/30)      2. ["A Dataset of Peer Reviews (PeerRead): Collection, Insights and NLP Applications"](https://github.com/allenai/PeerRead)   2. **검증 데이터**      1. 특허청 open API를 활용하여 특허 명세서, 요약서 등      2. Google scholar      3. PatentsView(USPTO)      4. [Ground AI](https://www.groundai.com/) 에서 크롤링/API |

1. 국내 특허(실용신안) 분쟁 수는 연간 500건 이상 발생(특허법원, 연도별 사건통계, 각연도.) [↑](#footnote-ref-0)
2. 잘 알려지지 않은 특허를 활용하여 소송을 제기하는 특허 사냥꾼(patent shark)의 활동으로 인해 반도체 제조사인 인텔이 6억 7500만 달러를 지불하는 등의 피해 사례도 다수 존재함(Henkel, J., Reitzig, M. Patent sharks. *Harvard Business Review*, 2008, pp.129-133.) [↑](#footnote-ref-1)