

1. 개요: 논문 요약 데이터셋이란?

텍스트 요약은 컴퓨터 프로그램을 이용해 문서의 축약된 내용을 자동으로 생성하는 것을 의미하며 딥러닝에 기반한 자연어 이해와 생성 기술이 필요하다. 논문 요약의 경우 기존의 추출요약 방식이 아닌 생성요약 방식으로 구축하였으며 OA논문(18만 건)과 특허명세서(17만 건)를 대상으로 요약대상 원문과 요약문이 쌍으로 구성될 수 있도록 하였다. 학술논문은 한국학술지인용색인(KCI), 국립중앙도서관, KISTI 등의 관련 연구기관에서 제공하는 Open Access 한국어 학술 논문을 대상으로 수집하여 논문의 전체(초록)와 섹션(논문 원문 중 섹션 일부 Paragraph)에 해당하는 요약문을 작성, 활용서비스 데이터 제공하고 특허자료의 경우 한국특허정보원에서 제공하는 특허명세서를 구매하여 전체 요약과 섹션(본문 중 섹션 일부 Paragraph)에 해당하는 요약문을 작성하고, 활용서비스용 데이터를 가공하였다.

데이터의 수집과 정제, 가공은 (주)나라지식정보와 (주)단아코퍼레이션이 공동으로 수행하였고 데이터 모델링과 활용서비스 개발은 (주)포티투마루에서 담당하였다.

현재까지는 제목이나 초록, 저자 등 키워드만으로 검색해 정형화된 정보를 얻을 수밖에 없고 관심 있는 논문의 연구 주제나 연구 방법, 연구 결과를 파악하려면 일일이 원문을 직접 읽어보고 판단해야 했지만 논문 요약 데이터셋 구성과 활용 서비스를 이용하여 AI가 논문을 대신 읽고 요약해주는 서비스 개발을 목표로 활발한 연구와 다양한 서비스 개발을 진행하고 있다.

기계 독해 기술은 질의응답 서비스에 주로 사용되며, 사람이 맥락을 이해하고 논리적으로 답을 찾는 것처럼 질의에 대한 답을 찾는 기술이며, AI 챗봇 상담, 방대한 전문지식에 대한 질의응답 및 시맨틱 검색 등에 활용하기 위해 활발한 연구 및 상용화를 진행하고 있다.

요약데이터의 원문과 라벨링 결과는 아래표를 참고

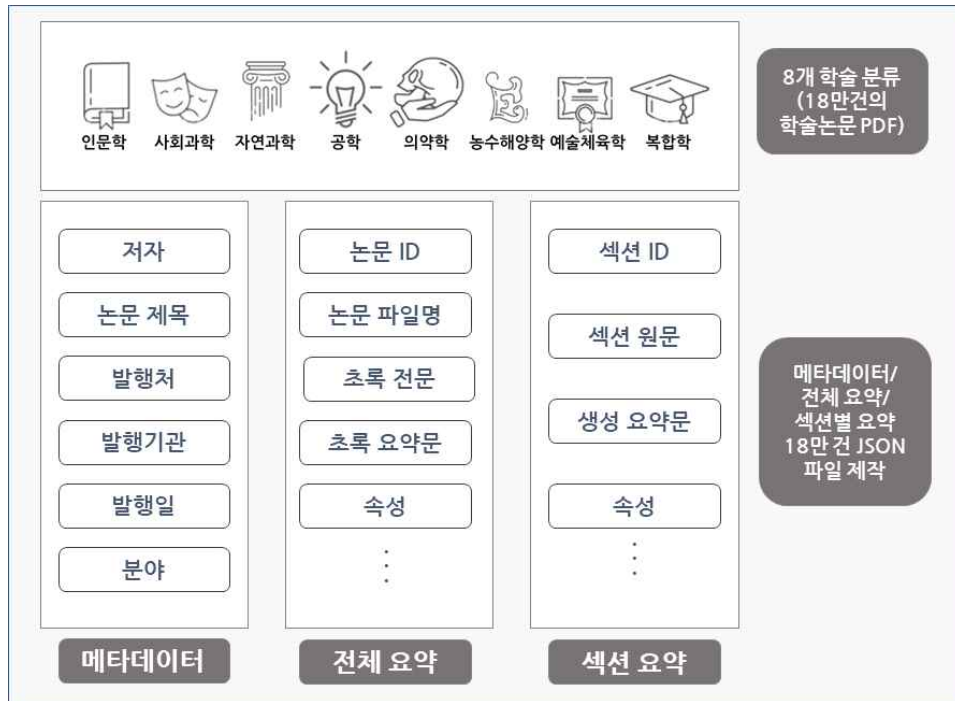
유형	예시	데이터 항목	json형식
이미지 (학술논문)		논문 원본 PDF	<pre>{ "doc_id": "A201008022769", "title": "국제 항공 여객운송의 네트워크 구조 분석", "date": "201711", "reg_no": "ART002281403", "ipc": "공학", "issued_by": "한국항공경영학회", "author": "김벽진 /Byeokjin Kim (제1) 안승범 /Seung-Bum Ahn (교신)", "summary_entire": [{ "original_text": "이 논문은 1995-2015년의 20년 동안 국제 항공 여객운송 네트워크의 위상적 특성과 진화 과정을 분석하고자 하였다. 중심성의 분포 및 구조적 특성과 같은 네트워크의 다양한 차원에서 시스템의 변화를 관찰하였다. 먼저 밀도,</pre>

	<div>한국항공경영학회지 제15권 제5호, 2017년 10월</div> <div>국제 항공 여객운송의 네트워크 구조 분석*</div> <div>김백진* · 안승범**</div> <div>Analysis on the Structure of International Passenger Air Transportation Network</div> <div>Byeok-Jin Kim · Seung-Bum Ahn</div> <div>국문초록</div> <div><p>이 논문은 1995-2015년의 20년 동안 국제 항공 여객운송 네트워크의 위상적 특성과 진화 과정을 분석하고자 하였다. 중심성의 분포 및 구조적 특성과 같은 네트워크의 다양한 차원에서 시스템의 변화를 관찰하였다. 먼저 밀도, 연결집중도, 매개집중도, 클러스터링 구조 등 일련의 측정 결과를 제시하였다. 또한 공항 시스템의 경로 특성 및 네트워크 효율성에 대해서도 조사하였다. 네트워크 시스템의 역동성을 포착하기 위해 측정치 간의 상관관계를 탐구하고 네트워크 연결체계 내에서 노드의 상호작용을 조사하였다.</p><p>분석 결과 국제운송 네트워크는 장기적으로 허브 집중형 네트워크에서 분산형으로 이행하고 있고, 네트워크 효율성은 노드 수의 증가에도 불구하고 안정된 추이를 보이고 있다. 네트워크 연결체계 면에서는 유럽과 미주의 도시들이 London으로의 집중 현상이 뚜렷이 나타난 반면, 아시아·태평양의 도시들은 분산 현상이 나타났다. 1995년의 네트워크 연결체계는 London, New York, Tokyo의 글로벌 3분 체계였으나, 2015년에는 London, Hong Kong, Dubai가 네트워크 권역을 주도하였다.</p></div> <div>주제어 : 항공운송, 네트워크, 연결정도, 매개중심성, 클러스터링, 컴포넌트</div> <div><p>* 인천대학교 동북아문명대학원 박사과정, 주저자, akim@jin.naver.com ** 인천대학교 동북아문명대학원 교수, 교신저자, sbahn@incheon.ac.kr</p></div>		<p>연결집중도, 매개집중도, 클러스터링 구조 등 일련의 측정 결과를 제시하였다. 또한 공항 시스템의 경로 특성 및 네트워크 효율성에 대해서도 조사하였다. 네트워크 시스템의 역동성을 포착하기 위해 측정치 간의 상관관계를 탐구하고 네트워크 연결체계 내에서 노드의 상호작용을 조사하였다. 분석 결과 국제운송 네트워크는 장기적으로 허브 집중형 네트워크에서 분산형으로 이행하고 있고, 네트워크 효율성은 노드 수의 증가에도 불구하고 안정된 추이를 보이고 있다. 네트워크 연결체계 면에서는 유럽과 미주의 도시들이 London으로의 집중 현상이 뚜렷이 나타난 반면, 아시아·태평양의 도시들은 분산 현상이 나타났다. 1995년의 네트워크 연결체계는 London, New York, Tokyo의 글로벌 3분체계였으나, 2015년에는 London, Hong Kong, Dubai가 네트워크 권역을 주도하였다.",</p> <p>"summary_text": "이 논문은 1995-2015년 동안 국제 항공 여객운송 네트워크의 위상적 특성 및 진화 과정을 분석하였다. 관찰한 바 네트워크 연결체계 면에서 유럽과 미주의 도시들이 London으로의 집중 현상이 나타난 반면, 아시아·태평양의 도시들은 분산 현상이 나타났다."</p> <p>}</p> <p>},</p> <p>"summary_section": [</p> <p>{</p> <p>"orginal_text": "항공운송은 국가의 경제, 사회 발전에 중요한 역할을 담당하는 기간 시스템이다. 경제가 급속하게 성장하고 세계화 과정이 가속화되면서 더 많은 사람들이 다양한 요구를 실현하기 위해 항공 운송을 선택하고 있다. 따라서 항공운송 기간 시스템의 발전 과정을 이해하는 것이 이전보다 훨씬 중요해지고 있다. 항공운송산업은 규제 완화와 세계 경제의 글로벌화로 인해 항공운송의 구조가 선형 노선(linear route) 시스템에서 네트워크(network) 시스템으로 전환되고 있다(신태진, 노태우, 이윤철, 2016). 운송시장에서 네트워크란 한정된 자원을 이용하여 상품을 생산, 판매하기 위해 구성되는 노선의 집합을 의미한다. 네트워크는 경제성이 있는 각 개별 노선의 유기적인 결합체로서, 운송지점의 수와 각 지점의 스케줄을 결정함으로써 하나의 시스템으로 운영된다(허희영, 2002) 항공사가 네트워크의 여러 노선들에 투입할 항공기 기종을 결정할때로 사용되어야 한다. 항공기의 기종은 택인 바를 수는 없다. 운항승무원 훈련, 정비, 지상설비에 대한 투자로 인해 항공기 기종은 고정되는 특성이 있고 수년 후에나 변경할 수 있다. 항공사가 운항을 위해 항공기를 도입하면 운항비용에 영향을 미치는 가장 중요한 요인은 투입 비용의 수준을 제외하고는 항공사가 운항할 노선구조이다(Doganis, 2010).</p> <p>따라서 항공사의 신규 노선 개설, 기존 노선의</p>
텍스트 (메타데이터)	<div>한국항공경영학회지 제15권 제5호, 2017년 10월</div> <div>국제 항공 여객운송의 네트워크 구조 분석*</div> <div>김백진* · 안승범**</div> <div>Analysis on the Structure of International Passenger Air Transportation Network</div> <div>Byeok-Jin Kim · Seung-Bum Ahn</div> <div>국문초록</div> <div><p>이 논문은 1995-2015년의 20년 동안 국제 항공 여객운송 네트워크의 위상적 특성과 진화 과정을 분석하고자 하였다. 중심성의 분포 및 구조적 특성과 같은 네트워크의 다양한 차원에서 시스템의 변화를 관찰하였다. 먼저 밀도, 연결집중도, 매개집중도, 클러스터링 구조 등 일련의 측정 결과를 제시하였다. 또한 공항 시스템의 경로 특성 및 네트워크 효율성에 대해서도 조사하였다. 네트워크 시스템의 역동성을 포착하기 위해 측정치 간의 상관관계를 탐구하고 네트워크 연결체계 내에서 노드의 상호작용을 조사하였다.</p><p>분석 결과 국제운송 네트워크는 장기적으로 허브 집중형 네트워크에서 분산형으로 이행하고 있고, 네트워크 효율성은 노드 수의 증가에도 불구하고 안정된 추이를 보이고 있다. 네트워크 연결체계 면에서는 유럽과 미주의 도시들이 London으로의 집중 현상이 뚜렷이 나타난 반면, 아시아·태평양의 도시들은 분산 현상이 나타났다. 1995년의 네트워크 연결체계는 London, New York, Tokyo의 글로벌 3분 체계였으나, 2015년에는 London, Hong Kong, Dubai가 네트워크 권역을 주도하였다.</p></div>	출처 (발행처) 발행일 논문제목 저자	
텍스트 (초록)	<div>국문초록</div> <div><p>이 논문은 1995-2015년의 20년 동안 국제 항공 여객운송 네트워크의 위상적 특성과 진화 과정을 분석하고자 하였다. 중심성의 분포 및 구조적 특성과 같은 네트워크의 다양한 차원에서 시스템의 변화를 관찰하였다. 먼저 밀도, 연결집중도, 매개집중도, 클러스터링 구조 등 일련의 측정 결과를 제시하였다. 또한 공항 시스템의 경로 특성 및 네트워크 효율성에 대해서도 조사하였다. 네트워크 시스템의 역동성을 포착하기 위해 측정치 간의 상관관계를 탐구하고 네트워크 연결체계 내에서 노드의 상호작용을 조사하였다.</p><p>분석 결과 국제운송 네트워크는 장기적으로 허브 집중형 네트워크에서 분산형으로 이행하고 있고, 네트워크 효율성은 노드 수의 증가에도 불구하고 안정된 추이를 보이고 있다. 네트워크 연결체계 면에서는 유럽과 미주의 도시들이 London으로의 집중 현상이 뚜렷이 나타난 반면, 아시아·태평양의 도시들은 분산 현상이 나타났다. 1995년의 네트워크 연결체계는 London, New York, Tokyo의 글로벌 3분 체계였으나, 2015년에는 London, Hong Kong, Dubai가 네트워크 권역을 주도하였다.</p></div>	논문 전체 요약	
텍스트 (섹션)	<div>I. 서론</div> <div><p>항공운송은 국가의 경제, 사회 발전에 중요한 역할을 담당하는 기간 시스템이다. 경제가 급속하게 성장하고 세계화 과정이 가속화되면서 더 많은 사람들이 다양한 요구를 실현하기 위해 항공 운송을 선택하고 있다. 따라서 항공운송 기간 시스템의 발전 과정을 이해하는 것이 이전보다 훨씬 중요해지고 있다.</p><p>항공운송산업은 규제 완화와 세계 경제의 글로벌화로 인해 항공운송의 구조가 선형 노선(linear route) 시스템에서 네트워크(network) 시스템으로 전환되고 있다(신태진·노태우·이윤철, 2016).</p><p>운송시장에서 네트워크란 한정된 자원을 이용하여 상품을 생산, 판매하기 위해 구성되는 노선의 집합을 의미한다. 네트워크는 경제성이 있는 각 개별 노선의 유기적인 결합체로서, 운송지점의 수와 각 지점의 스케줄을 결정함으로써 하나의 시스템으로 운영된다(허희영, 2002).</p><p>항공사가 네트워크의 여러 노선들에 투입할 항공기 기종을 결정할때로 사용되어야 한다. 항공기의 기종은 택인 바를 수는 없다. 운항승무원 훈련, 정비, 지상설비에 대한 투자로 인해 항공기 기종은 고정되는 특성이 있고 수년 후에나 변경할 수 있다. 항공사가 운항을 위해 항공기를 도입하면 운항비용에 영향을 미치는 가장 중요한 요인은 투입 비용의 수준을 제외하고는 항공사가 운항할 노선구조이다(Doganis, 2010).</p><p>따라서 항공사의 신규 노선 개설, 기존 노선의</p></div>	논문 섹션별 요약	

그림 1 논문 요약의 사례

2. 데이터셋의 구성

○ OA 논문



○ 특허명세서



○ 어노테이션 포맷

구분	요소명		예시	유형	길이	필수여부	설명
	한글	영문명					
메타 데이터	아이디	doc_id	A201008022769	string		Y	논문 1편을 기준으로 한 id
	논문/특허명	title	국제 항공 여객운송의 네트워크 구조 분석	string		Y	
	발행/등록일자	date	201711	string		Y	
	파일명	reg_no	ART002281403	string		N	논문 원문(PDF) 파일명
	한국연구재단/특허 분류정보	ipc(class)	공학(특허의 경우 ipc 분류 적용)	string		Y	- 한국연구재단 (nrf.re.kr)의 한국학술지인용색인(KCI)의 연구분야 분류체계 적용 -특허의 경우 ipc 분류 적용
	발행기관	issued_by	한국항공경영학회	string		Y	
	저자	author	김벽진 /Byeokjin Kim (제1)안승범 /Seung-Bum Ahn (교신)	string		Y	
전체 요약 및 섹션 요약	전체 요약 구분	summary_entire		string		Y	전체 요약 부분
	전체 원문 (초록)	original_text	이 논문은 1995-2015년의 20년 동안 국제 항공 여객운송 네트워크의 위상적 특성과 진화 과정을 ... (중략) ...네트워크 연결 체계 면에서는 유럽과 미국의 도시들이 London으로의 집중 현상이 뚜렷이 나타난 반면, 아시아-태평양의 도시들은 분산 현상이 나타났다. 1995년의 네트워크 연결체계는 London, New York, Tokyo의 글로벌 3분체계였으나, 2015년에는 London, Hong Kong, Dubai가 네트워크 권역을 주도하였다.	string		Y	논문 초록 전문
	전체 요약문	summary_text	이 논문은 1995-2015년 동안 국제 항공 여객운송 네트워크의 위상적 특성 ... (중략) ... 아시아-태평양의 도시들은 분산 현상이 나타났다.	string		Y	논문 초록 요약문
	섹션 요약 구분	summary_section		string		Y	섹션 요약 부분
	섹션 단락	original_text	항공운송은 국가의 경제, 사회 발전에 중요한 역할을 담당하는 기간 시스템이다. 경제가 급속하게 성장하고...(중략)...그러한 항공기의 기종은 결정된대로 사용되어야 한다.	string		Y	섹션 단락
	섹션 요약문	summary_text	경제 성장과 세계화로 많은 사람들이 항공 운송을 선택한다. 이 산업은 규제 완화와 세계 경제의 글로벌화로 구조가 네트워크 시스템으로 전환되었다. 이는 경제	string		Y	섹션 요약

특허 연도별 등록 건수			
2019	2018	계	구축 대상
125,661	119,012	244,673	170,000

4. 데이터 예시

■ 학술논문 요약 데이터 예시

```
{
  "totalcount": 32000,
  "data": [
    {
      "doc_type": "논문",
      "doc_id": "A201008176694",
      "title": "교직 선택 결정에 대한 후회 감정의 탐색적 해석",
      "date": "2018",
      "reg_no": "ART002387022",
      "ipc": "사회과학",
      "issued_by": "사범대학부속중등교육연구소",
      "author": "조석훈 /JO SEOG HUN (단독)",
      "summary_entire": [
        {
          "original_text": "본 연구는 경제협력개발기구(Organization for Economic Cooperation and Development, OECD)가 실시한 교수-학습 국제조사(Teaching and Learning International Survey 2013, TALIS) 중 교직 선택에 대한 교사 후회에 관한조사를 연구의 출발점으로 삼았다. 한국의 경우 중학교 교사가 참여한 이 조사에서 응답 교사의 20.1%가 교사가 된 것을 후회한다는 결과가 나오자 사회적 관심이 집중되었다. 교직 선택을 후회하는 교사 비율은 조사대상국 중 최고로 높은 수치로 이는 곧 교사의 낮은 사기와 만족도를 표현하는 것으로 해석되었고 교원 정책에 대한 비판의 근거로 제시되었다. 이 연구는 교사가 경험하는 후회의 감정은 매우 복잡한 현상이고 다양한 변인이 작용하는 심리적메커니즘을 가지고 있다는 선행연구에 기초하여 이러한 단순한 해석이 갖는위험을 지적하고 탐색적인 차원에서 대안적인 해석의 가능성을 제시하였다. 기회(대안)와 후회, 정상(norm)과 후회, 후회의 기능이라는 3가지 영역을검토하면서, 교사 후회 현상은 다양한 변인의 영향을 받을 수 있음을 지적하고 관련 정책적 시사점도 아울러 모색하였다. 특히, 교사 후회는 교직 선택동기에 관한 연구, 교직을 선택한 시기에 관한 조사, 교직 선택이 당연한 집단이 가진 특성, 후회를 감소시키기 위한 교사의 교직 행동 등과 관련지어새롭게 해석될 수
```

있음을 보여주었다. 앞으로 후회 감정에 관련된 심리적 비교에서 고려되는 요소, 행위/비행위 사이 차이, 후회에 관련된 심리적 처리의 휴리스틱과 편향 등 후속 연구가 적절히 이루어질 때 교사 후회 현상에 대한 제대로 된 해석이 가능하고, 교사 후회 현상을 바탕으로 한 정책 개발 및 평가도 타당하게 이루어질 수 있을 것이다.",

"summary_text": "본 논문은 경제협력개발기구가 실시한 교수-학습 국제조사 가운데 교직 선택에 대한 교사 후회에 대한 조사를 연구의 출발점으로 하였다. 한국의 경우 중학교 교사가 참여한 이 조사에서 응답 교사의 20.1%가 교사가 된 것을 후회한다는 결과가 나와 사회적 관심을 받았다."

}

],

"summary_section": [

{

"original_text": "한국교총은 '교사가 된 것을 후회한다.'는 교사 비율이 우리의 경우 20.1%로 조사대상국 중 가장 높다고 지적하면서 교사의 사기를 진작할 특단의 대책을 마련할 것을 촉구하였다(한국교총, 2015a). 허주 외 연구(2015: 131-132)도 TALIS 2013결과를 해석하면서 상대적으로 높은 교사 후회 비율을 근거로 교사의 직무만족도가 매우 낮은 수준이라고 진단하였다. 한국교총은 1995년 5.31교육개혁을 비롯하여 학생인권조례 제정, 학부모 민원 등 교권 추락 가속화 등이 작용한 결과라고 보고 기존 교원 정책에 대한 비판적 관점을 취하였다(한국교총, 2015a). 한국교총이 지적한 대로 20%를 상회하는 교사가 교직 선택을 후회하면서 심리적 고통을 느끼는 감정 상태에 있다면(Landman, 1993: 4참조), 현재와 미래에 교사의 직무 전념과 매진에 문제가 있을 것으로 예상할 수 있다. 감정으로써 후회는 앞으로 전개될 행동의 안내자로서 과거와 미래를 연결하는 다리가 되기 때문이다(Zeelenberg & Pieters, 2007: 4).",

"summary_text": "한국교총은 교사의 사기를 진작할 특단의 대책을 마련할 것을 촉구하였다. 한국교총은 학생인권조례 제정, 학부모 민원 등 교권 추락 가속화 등이 작용한 결과라고 보고 기존 교원 정책에 대한 비판적 관점을 취하였다."

}

]

■ 특허명세서 요약 데이터 예시

```
{
  "totalcount": 38000,
  "data": [
    {
      "doc_type": "특허",
      "doc_id": "A201007128452",
      "title": "아우로라 키나아제의 억제에 의해 유사분열 진행을 억제하기 위한 화합물 및 그 화합물을 포함하는 약학적 조성물",
      "date": "2005",
```

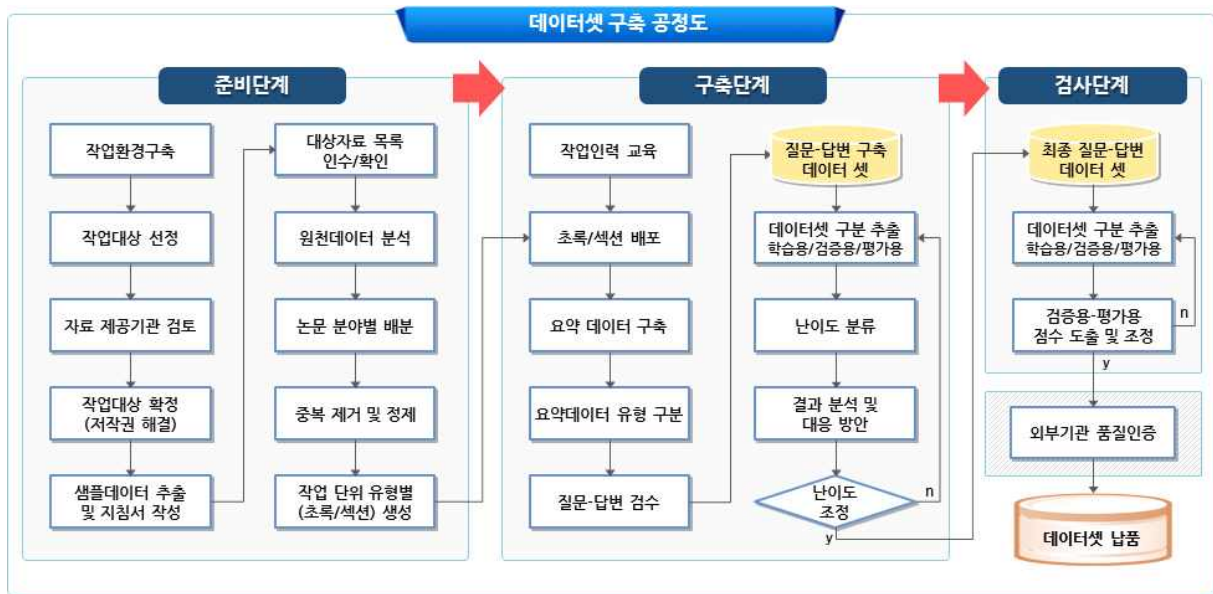
```

"reg_no": "1020067026332",
"ipc": "C07D 487/04(2006.01.01) C07D 487/14(2006.01.01) C07D 491/14(2006.01.01) A61K 31/55(2006.01.01)",
"author": "",
"summary_section": [
  {
    "original_text": "미국 암 학회(American Cancer Society)에 따르면, 2004년에는 대략 천사백만 명의 미국인들이 암으로 새로 진단받았으며, 대략 56만 명이 암으로 사망한 것으로 추산되었다. 의학의 진보가 암 생존율을 개선시키고는 있지만, 신규하고 더욱 효과적인 치료법이 계속적으로 요구되고 있는 실정이다. 암은 조절되지 않은 세포 증식에 의해 특징된다. 유사분열은, 일련의 복합 사건에 의해 크로모솜이 2개의 완전히 똑같은 딸세포로 분리되게 하는 세포 주기에서의 단계이다. 탁산 및 빈카 알칼로이드를 포함하는 현재의 다수 암 치료법은, 유사분열 조직을 억제시키도록 작용한다. 유사분열 진행은 주로, 유사분열 키나아제에 의해 매개되는 인산화 및 단백질 분해에 의해 조절된다. 아우로라 키나아제 부류의 구성원들(예를 들어, 아우로라 A, 아우로라 B, 아우로라 C)은, 센트로솜 분리, 스피indle 동역학, 스피indle 어셈블리 체크포인트, 크로모솜 정렬, 및 세포질분열을 조절함으로써 유사분열 진행을 조절한다 [참조: Dutertre et al., Oncogene, 21: 6175 (2002); Berdnik et al., Curr. Biol., 12: 640 (2002)]. 아우로라 키나아제의 과발현 및/또는 증폭은 결장 및 유방 종양을 포함하는 다수의 종양 유형에서 종양형성과 관련되어 있었다 [참조: Warner et al., Mol. Cancer Ther., 2: 589 (2003); Bischoff et al., EMBO, 17: 3062 (1998); Sen et al., Cancer Res., 94: 1320 (2002)]. 또한, 종양 세포에서 아우라 키나아제를 억제하면 유사분열 정지 및 세포자멸사가 야기되는 데, 이는 이들 키나아제가 암 치료를 위한 중요한 표적임을 제안한다 [참조: Ditchfield, J. Cell Biol., 161: 267 (2003); Harrington et al., Nature Med., 1 (2004)]. 사실상 모든 악성종양의 진행에서 유사분열이 중요한 역할을 담당하고 있기 때문에, 아우로라 키나아제의 억제제는 광범위한 사람 종양에 걸쳐 적용될 것으로 예측된다. 따라서, 신규한 아우로라 키나아제 억제제가 요구되고 있다.",
    "summary_text": "본 발명은 암 치료를 위한 화합물 및 그 방법에 관한 것이다. 구체적으로, 본 발명은 아우로라 키나아제 효소를 억제하는 화합물, 그 화합물을 포함하는 약제 조성물, 및 암 치료를 위한 그 화합물의 사용 방법을 제공한다."
  }
],
},

```

5. 데이터 구축 과정

■ 데이터 구축 개요

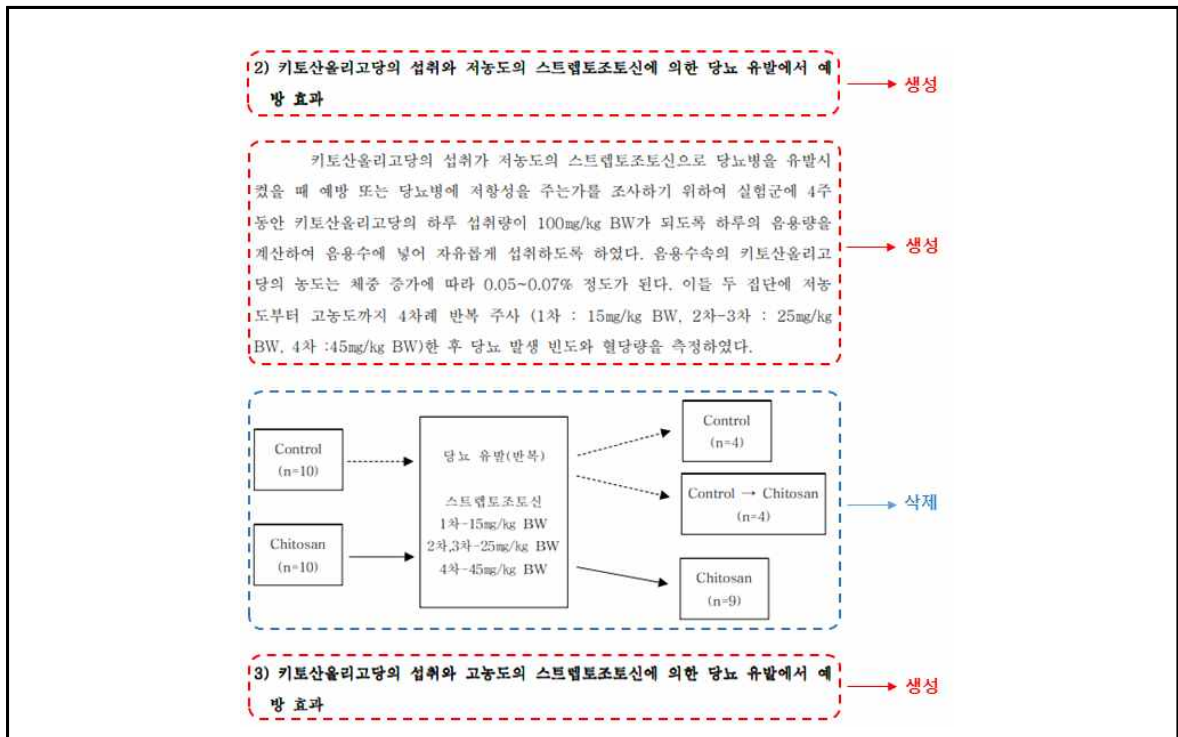


■ 준비단계 (수집 및 정제)

- Open Access, KISTI, 특허정보원 등에서 제공하는 원천데이터를 설정한 데이터 분류 체계를 바탕으로 수집
- 다양한 유형 및 형식을 보유한 논문제공기관의 데이터에 대하여 그림이나 표와 같이 요약으로 표현하기 어렵거나 불필요한 부분에 대한 정제 및 삭제를 통하여 데이터 전처리 작업 진행
- 논문 전체 요약, 색선별 요약을 위한 구분 작업 진행

○ 원천 데이터 정제 기준

- 문서 타이틀, 저자/발명가, 문서종류, 발행처 등의 필드를 구분하여 저장하고 데이터 정제를 진행하여 문서 내 불필요한 정보의 삭제 처리
- 논문 요약문 추출에 부적합한 그림, 서식 등의 삭제
- PDF에서 추출된 텍스트의 각종 오류 정제 후 데이터 구축



[논문의 비(非)텍스트 요소 제거 예시]

연구 목적: 이 연구는 Br Waa nemark 임플란트의 후향적인 연구를 통해 machined 임플란트의 15년, TiUnite TM 임플란트의 5년 누적 생존율을 구하여 비교하고, 위험 요소와의 상관관계를 밝히고자 하였다. **연구 재료 및 방법:** 본 연구는 1993년부터 2008년까지 고려대 구로병원 임플란트클리닉에서 임플란트 치료를 받은 환자를 대상으로 하여, 환자의 임상기록을 토대로 조사하였고, 임플란트 일차수술 후 즉시 부하를 가한 경우와 임플란트 보철물이 다른 임플란트 시스템과 결합한 경우는 제외하였다. 15년 간 총 155명의 환자에게 541개의 Br Waa nemark 임플란트 중, 264개의 machined 임플란트와 277개의 TiUnite TM 임플란트를 식립하였고, 임플란트 수술 및 보철수복 술식은 Adell 등이 제안한 방법에 따랐다. 6개월에서 1년 간격으로 Follow-up을 위한 재내원 기간 동안 임플란트에 대한 임상검사를 실시하였다. 본 연구의 관찰 기간은 1차 수술일로부터 2008년 12월 이내의 최종 내원일까지였고, 임상검사시 동요도, 타진, screw loosening, 그리고 환자의 불편감 등을 관찰하고, Zarb와 Albreksson에 의한 임플란트 생존 범주를 근거로 누적 생존율을 평가하였다. 임플란트 누적생존율(CSR)이 Kaplan Meier estimate를 이용하여 얻었으며, 각각의 위험요소가 누적생존율에 미치는 영향을 Cox proportional hazards regression을 이용하여 분석하였다 ($\alpha = .05$). **결과:** 전체 Br Waa nemark 임플란트의 15년 누적생존율은 86.07%이었고, machined 임플란트의 경우 15년 누적생존율은 82.89%, 5년 누적생존율은 89.21%였으나 Tiunite TM 임플란트의 5년 누적생존율은 98.74%로 machined surface가 Tiunite TM 임플란트에 비해 실패율이 4.6배 정도 더 높았다. 전신질환을 가진 환자의 경우 실패율이 32% 더 증가하였고, 상악동 거상술 또는 골이식을 동반한 수술 등 부가적인 수술을 받은 경우 임플란트 실패율이 40% 더 증가하여 위험 요소와 임플란트의 생존율은 상관 관계가 있었다. Kennedy 분류에 따른 부분 무치악부의 위치에 따라 임플란트 실패율이 통계적으로 유의하게 달랐으며, 고구치와 구치부가구치부가 전치부에 비해 임플란트 실패율이 높게 나타났다. 결론: Br Waa nemark machined 임플란트와 TiUnite TM 임플란트는 모두 우수한 임상적 결과를 보였으나, TiUnite TM 임플란트가 더 성공적이며, 임플란트의 누적 생존율은 위험요소와 연관성이 있었다.

Tiunite^TM (첨자 표시)

소제목 (제거필요)

Res .05 (한글자동변환오류)

[학술논문의 서식 기호 등 텍스트 정제 예시]

■ 구축 단계 : 학습데이터 구축(논문)

- 인공지능 생성요약(Abstractive Summarization) 모델 구축을 위한 데이터셋 구축 작업
- 유형별로 분류 및 정제된 대상 논문 내용을 참고하여 논문 전체 및 섹션별로 요약문을 작성
- 인공지능 생성요약 알고리즘의 편향 방지를 위해 다양한 분야의 데이터를 구축대상으로 하여 구축하도록 하고, 각 데이터의 주제 분류, 논문저자, 날짜 등의 정보를 통해 골고루 구축되도록 구성함
- 생성 요약문의 길이는 요약대상 원문의 10~30% 이내로 요약하고 원천 데이터의 내용을 압축적으로 표현되도록 구축작업을 수행

■ 구축 단계 : 학습데이터 구축(특허명세서)

- 인공지능 생성요약(Abstractive Summarization) 모델 구축을 위한 데이터셋 구축 작업
- 특허명세서의 전체 요약문 및 섹션 요약문을 작성
- 각 데이터의 출원번호, 출원인, 등록번호 등의 정보를 통해 골고루 구축되도록 구성함
- 특허명세서 전체 원문 텍스트를 대상으로 요약문(300자 이내)을 작성하고 섹션요약은 특허명세서의 주요 섹션(발명의 방법, 발명의 원리 등)에 대한 섹션 요약을 실시

■ 검사단계

- 작업자와 별도로 검수자가 구축된 학습데이터를 검수하여 본 사업의 생성요약 모델 구축을 위한 학습데이터가 고품질을 유지하도록 수행
- 검수자가 데이터 구축 과정을 확인하고 불필요한 정보의 제거 및 별지 등의 정보가 누락 되지 않도록 처리
- 검수자가 데이터를 검수 완료하게 되면 생성요약 학습데이터로 최종 등록

6. 데이터 구축 담당자

참여기관 : (주)나라지식정보 (전화: 02-3141-7644), 이메일: nara@narainformation.com