

국내 최대 | 강력한 융합 데이터 | 소셜 API

VAIV COMPANY

Sometrend Data+ 서비스 소개서

Copyright © VAIV Company. All rights reserved.

CONTENTS.

01

VAIV Company

02

서비스 정보

03

서비스 특징점

04

데이터 상품

05

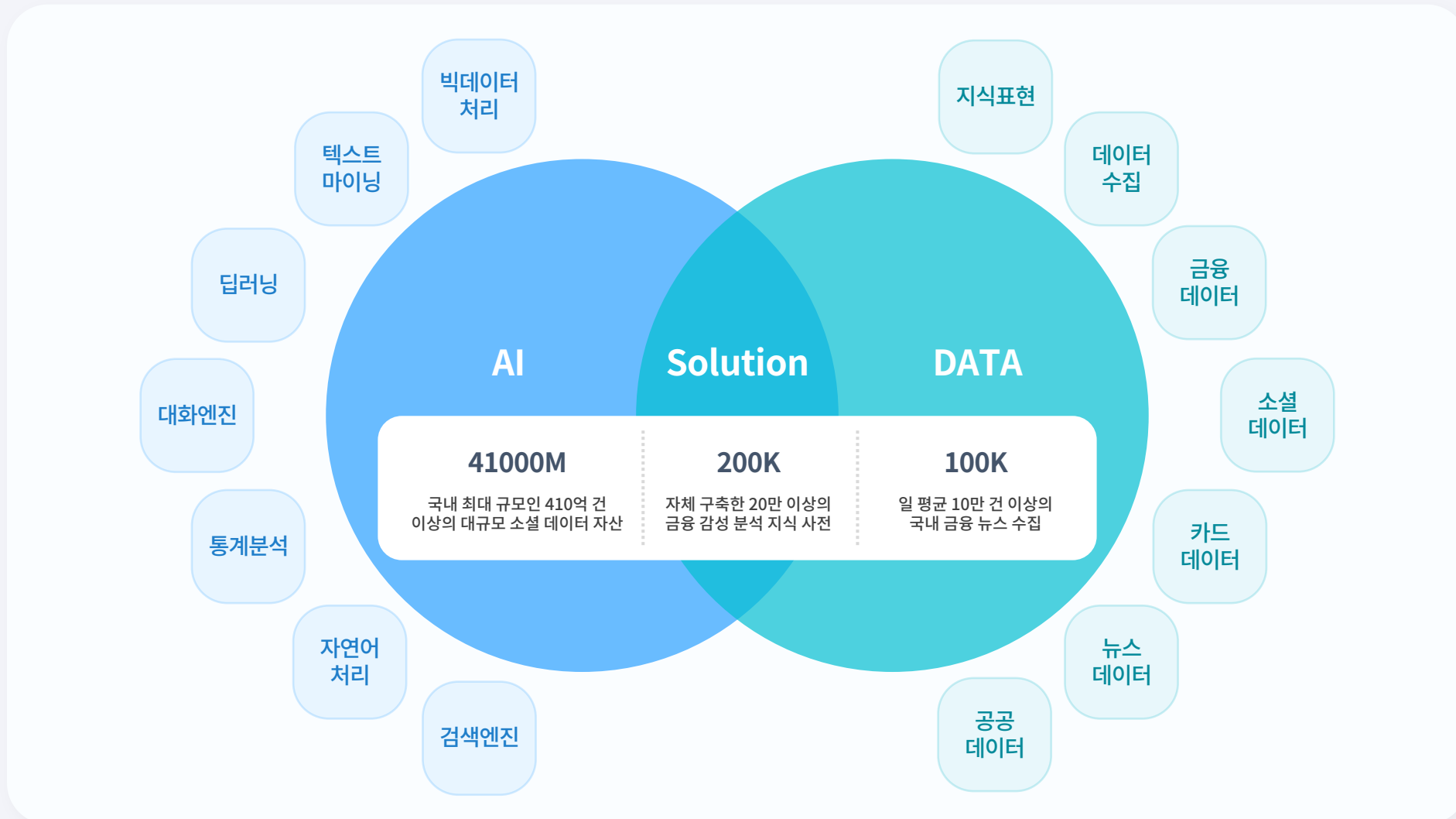
서비스 가격 정책

01. VAIV Company

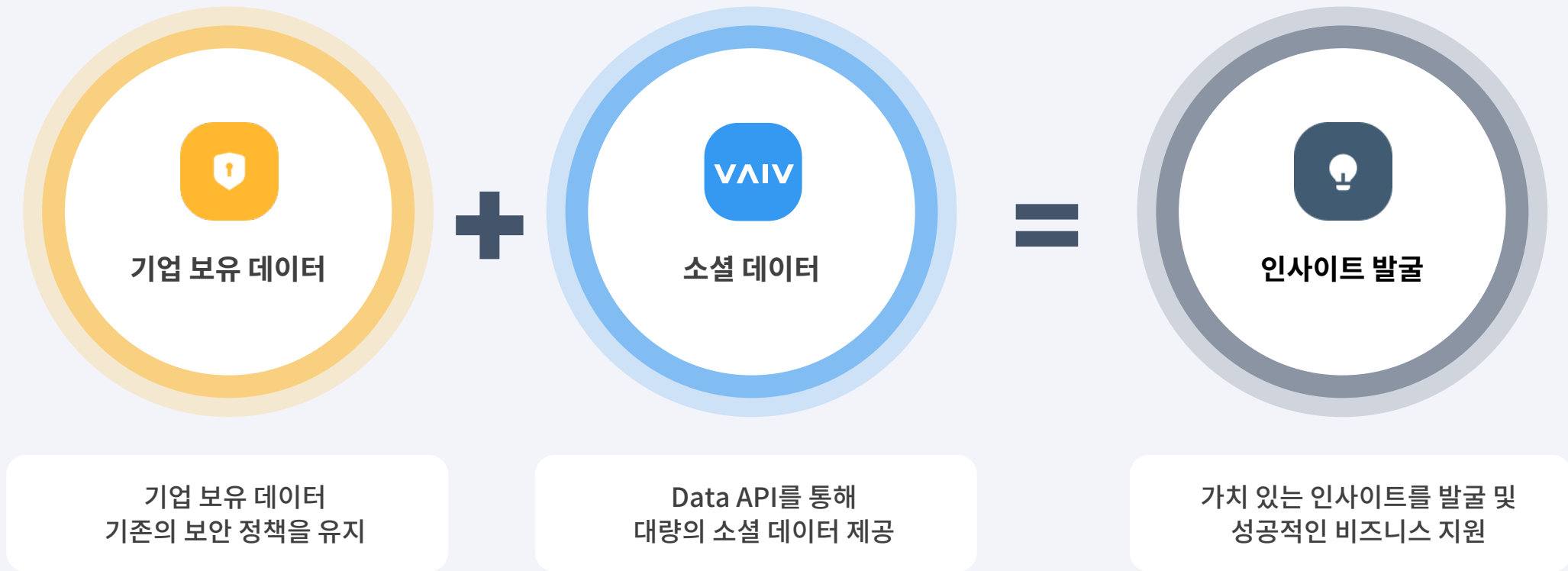
국내 최초의 AI 기반 빅데이터 분석 전문 기업 바이브컴퍼니 ('구 다음소프트' 2000년 7월 (주) 다음 커뮤니케이션에서 분사)
빅데이터 기술을 통해 더 나은 가치를 연구하는 기업, 바이브컴퍼니



20년간 축적된 데이터와 자연어 처리 인공지능 기술을 바탕으로 비즈니스의 혁신 추구



기업 내부 데이터와 바이브컴퍼니의 소셜 데이터 융합을 통해 인사이트 발굴과 의사결정 지원



02. 서비스 정보

서비스 정보

강력한 융합 데이터 Data+, 기업 내부 데이터와 소셜 빅데이터 융합을 통한 다양한 인사이트 발굴 지원



제공 API 정보

소셜 API

키워드 빈도

문서 목록

연관어

- 연관어(분류, 속성, 감성)

연관어 추이

- 연관어 추이(속성, 감성)

기간별 상위 연관어

- 기간별 상위(속성, 감성)

리뷰 API

기초 정보

- 카테고리 정보 조회
- 브랜드 정보 조회
- 제조사, 쇼핑몰 정보 조회

제품 상세정보

추이 분석

- 리뷰 추이
- 평점 추이
- 긍정률 추이
- 긍부정 키워드

랭킹 분석

감성 분석

기타 지원

금융 감성 API

금융감성 카테고리

일별 금융감성 데이터

금융감성 긍부정 추이

금융감성 뉴스 원문 정보

금융감성지수

재무제표 API

관심 종목/계정과목 조회

- 관심 종목 목록
- 관심 계좌과목 목록
- 관심 종목/계정과목 데이터
- 관심 종목/계정과목 신규 데이터

종목별 재무제표 조회







- 종목 검색
- 종목별 재무제표 종류
- 종목별 재무제표 데이터

주요 계좌과목 비교

- 주요 계좌과목
- 종목별 비교
- 계정별 비교
- 분기별 비교

수집 범위

최대 문서 수집을 위해 트위터, 블로그, 뉴스, 커뮤니티, 유튜브를 대상으로 채널 별 특성에 따라 계정 기반 또는 API를 활용하여 수집

 트위터	<ul style="list-style-type: none"> ▪ API 기반 수집 (트위터에서 제공하는 스트리밍 API를 사용하여 수집) ▪ 한국어를 포함한 트윗을 실시간 수집
 인스타그램	<ul style="list-style-type: none"> ▪ 팔로워 상위 30만 계정 문서 수집 ▪ 요청 시 추가 계정 수집 제공 가능 (팔로워수, 원문, 좋아요수, 댓글 수집 제공)
 뉴스	<ul style="list-style-type: none"> ▪ 뉴스 목록 페이지에서 새로운 뉴스 발견 시 페이지 방문하여 뉴스 수집 ▪ 네이버 포털 뉴스(네이버 제휴 주요 언론사) 에 등록되는 뉴스 실시간 수집 ▪ 뉴스 기사 별 댓글 수 수집
 블로그	<ul style="list-style-type: none"> ▪ 계정 기반 수집(약 1,500만 개의 네이버 블로그 계정에 대하여 문서 수집) ▪ 파워블로거를 우선으로 블로그 페이지를 1일 1회 방문하여 새로운 포스트를 수집 ▪ 새로운 블로그 계정은 방문한 블로그 페이지에서 이웃이나 친구 관계로 블로그 계정 추가
 커뮤니티	<ul style="list-style-type: none"> ▪ 게시판 별로 1일 1회 방문하여 새로운 게시글 수집 (1,700여개 커뮤니티, 7,800여개 게시판) ▪ 게시판, 네이버/다음 카페 등 일반적인 웹 게시물을 게시판 별로 설정에 기반하여 수집 ▪ 디시인사이드, 뽐뿌, 루리웹, 에펨코리아, 네이트판, 클리앙 등 대표 커뮤니티 사이트 수집
 유튜브	<ul style="list-style-type: none"> ▪ 구독자 1000명 이상 기준 채널 1일 1회 방문 수집 (1,140,000+ 계정) ▪ 요청 시 경쟁사 채널 등 추가 가능 (원문, 팔로워수, 좋아요수, 댓글 등)

수집 방식

사전에 선정한 키워드만 수집하는 방식이 아닌, 데이터 전수 수집으로 분석 키워드 제약 없이 즉시 활용 가능

NO

타사 키워드 수집 방식

포털이나 검색 시스템의 키워드 결과 수집 방식

- ! 포털이나 검색 시스템에서 조회되지 않는 페이지 수집 불가
- ! 소셜 분석을 원하는 키워드를 사전에 선정 후 포털이나 검색 시스템에서 수집을 시작하여 사전 작업 대기 시간 필요
- ! 새로운 이슈가 발생하여 키워드를 추가할 필요가 있을 때에도 사전 작업 시간 필요



YES

바이브컴퍼니 수집 기술

대용량 데이터 분산 처리

- ✓ 병렬·분산처리 기술을 통해 일 평균 1,300만 건 이상의 대용량 데이터를 안정적으로 수집
- ✓ 전수집을 목표로 수집원의 특성을 반영한 수집 시스템 개발 운영

최신성 및 정확성

- ✓ 효과적인 수집 및 자원 활용을 통해 시스템의 안정성과 빅데이터 분석 정보의 최신성 유지
- ✓ 소셜 분석 관점에서 가능한 많은 데이터 수집을 위해 노하우가 담긴 수집 기법 운영
- ✓ 분석 대상 키워드의 제약 없이, 키워드 분석을 위한 불필요한 대기 시간 제거

강력한 수집 후처리

- ✓ 수집된 데이터를 중복 및 유사 문서 처리, 스팸 필터링 등의 강력한 후처리 수행
- ✓ 정확한 이슈의 크기를 분석하기 위해 중복 문서와 유사 문서 제거
- ✓ 블로그 데이터의 경우, 다양한 기법으로 광고성 문서 처리
- ✓ 상위 과정을 통해 인사이트 발굴의 신뢰성을 확보 및 양질의 데이터 제공

효율적인 데이터 저장 체계

- ✓ 대용량 데이터의 실시간 분석이 가능하도록 바이브컴퍼니의 빅데이터 시스템에 실시간 저장
- ✓ 서비스 목적 및 활용 용도에 따라서 다양한 형태의 저장 체계를 지원

03. 서비스 특징점

1

국내 최고 수집 기술력 보유

- 목적성, 포괄성, 안정성을 담보한 국내 최다 데이터 보유
- 선제적 데이터 수집으로 즉시 데이터 제공 가능

2

텍스트 마이닝 분야 최고의 기술력 보유

- 20년간 축적해온 자연어 처리 기술 및 온톨로지 보유
- 200만 어휘 이상의 개체명 사전 및 분류체계 보유

3

빅데이터 플랫폼 관련 최고의 기술력 보유







- 300대 이상의 하둡 클러스터 자체적으로 운영
- 1PB 이상의 데이터 관리 기술 보유
- 임팔라, 하이브, H-Base 등의 하둡 관련 기술력 보유

4

다양한 빅데이터 관련 사업의 성공적 수행

- 주요 통신사에 소셜 분석 API 제공 사례 보유
- 보유수백 개의 ASP 이용 고객 및 마케팅 보고서 사례 보유

국내 최대 데이터 커버리지

채널	일간 수집 데이터량	총 데이터 보유량	데이터 보유 기간	제공 데이터	수집 및 분석 문서 범위	수집	분석 주기
 트위터	1,000만 ~ 1,500만 건	354억	2011년~	본문, 작성자, 게시일, RT수	한글로 된 전체 데이터 + RT 수 분석	실시간	2시간
 뉴스	2만 ~ 3만 건	1.2억	2008년~	제목, 본문, 작성자, 게시일, 댓글 수	네이버 포털 뉴스 + 댓글 수 분석	실시간	2시간
 인스타그램	10만~20만 건	24억	2014년~	제목, 본문, 해시태그, 작성자, 게시일, 좋아요 수	팔로워 상위 30만개 계정 본문 및 해시태그 데이터 분석	매일	2시간
 블로그	30만~40만 건	18억	2008년~	제목, 본문, 작성자, 게시일	네이버 블로그 분석	매일	1일
 커뮤니티/ 카페	15만~25만 건	3.8억	2013년~	커뮤니티 명, 게시판 명, 제목, 본문, 작성자, 게시일	대중적, 유명한 카페 및 커뮤니티의 약 7,000개 게시판 분석	매일	1일
 유튜브	2만~3만 건	0.32억	2018년~	제목, 내용, 카테고리, 댓글, 해시태그, 구독자 수, 조회수, 댓글 수, 좋아요 수	구독자 1,000명 이상의 채널 114만개 이상 분석 (동영상 제목, 상위 댓글, 댓글 수, 조회수 분석)	매일	1일

최장 데이터

2008년 데이터부터 14년 이상의 누적 데이터 보유

최대 데이터

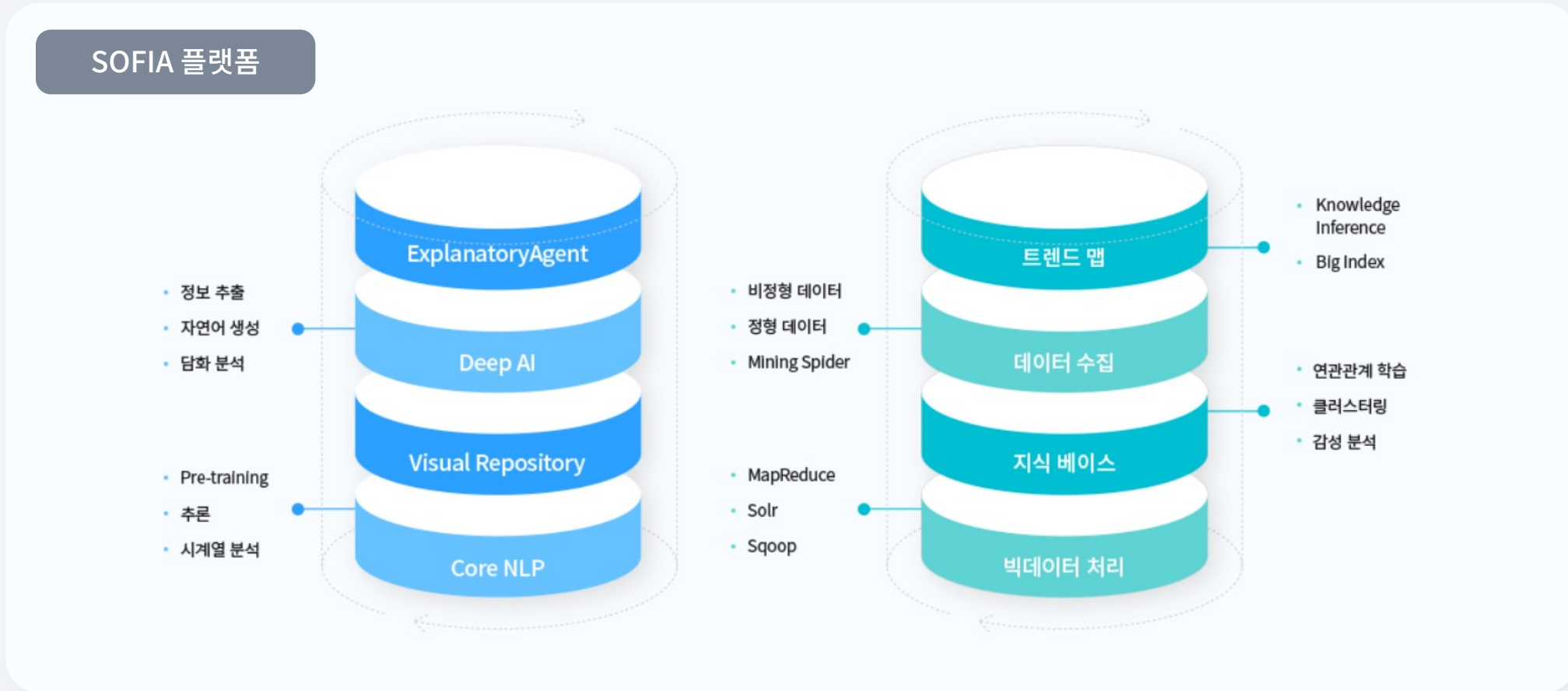
2022년 기준 410억 건

최다 채널

트위터, 인스타그램, 블로그, 커뮤니티/카페, 뉴스, 유튜브

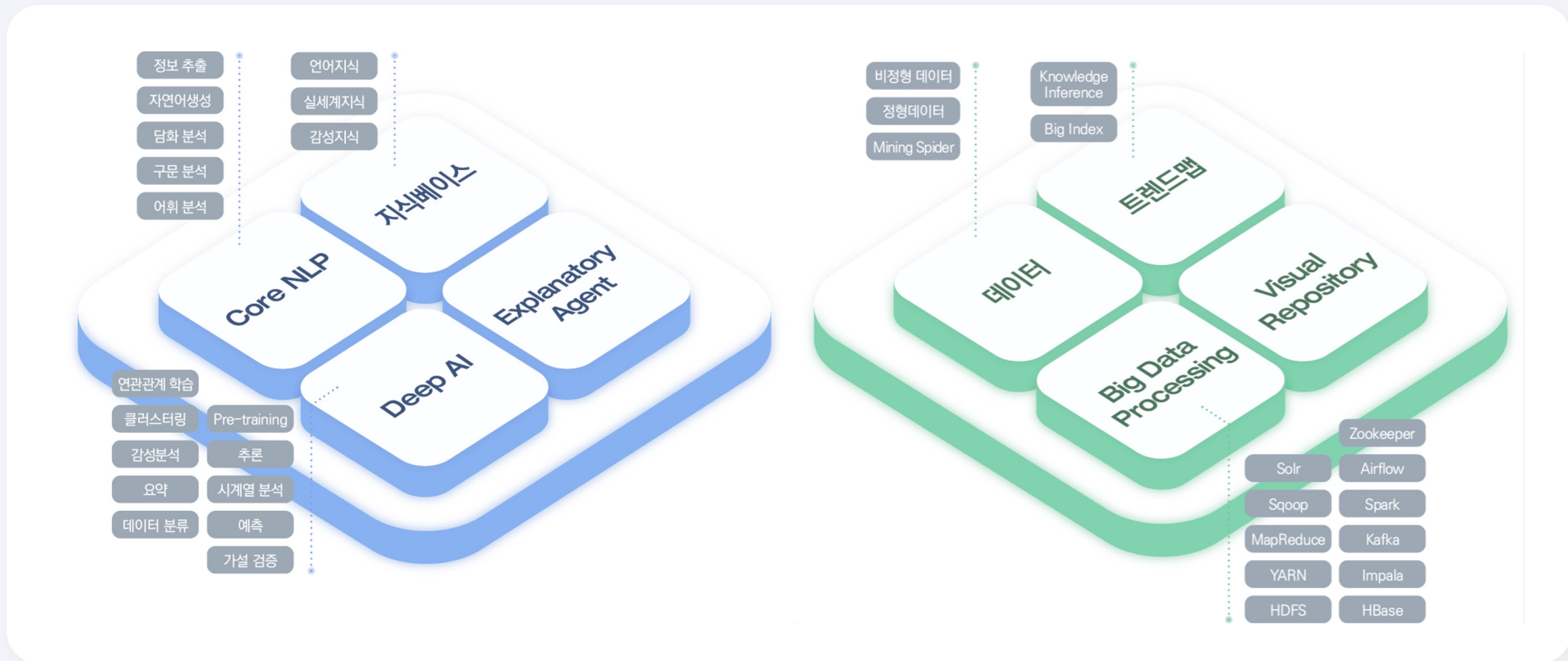
업계 최고 수준의 자연어 처리 기술

SOFIA 플랫폼을 통한 업계 최고 수준의 자연어 처리 기술



빅데이터 플랫폼 관련 최고의 기술력 보유

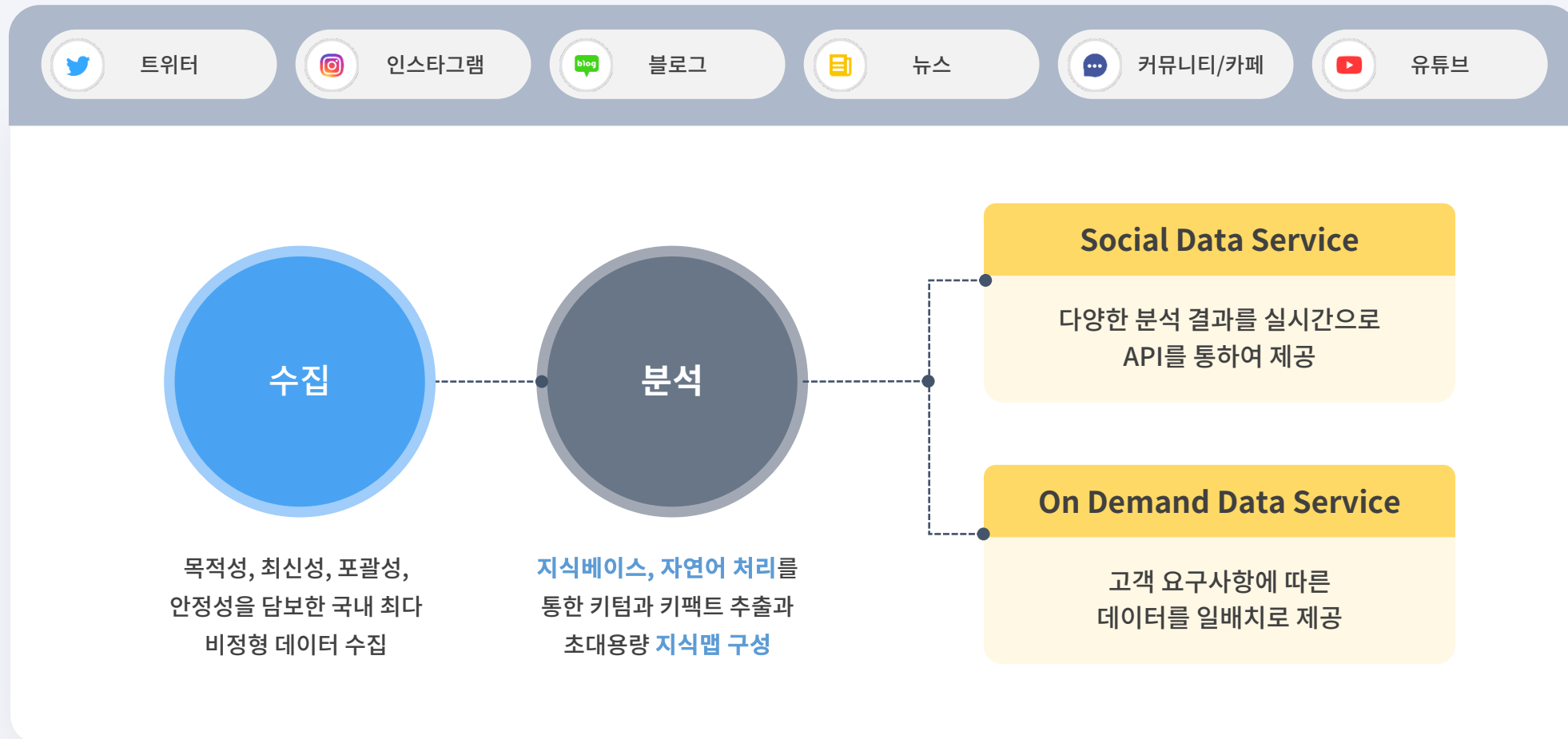
1PB 이상의 데이터 관리 기술 보유. 임팔라, 하이브, H-Base 등의 하둡 관련 기술력 보유





04. 데이터 상품

업계 최고 수준의 자연어 처리 기술

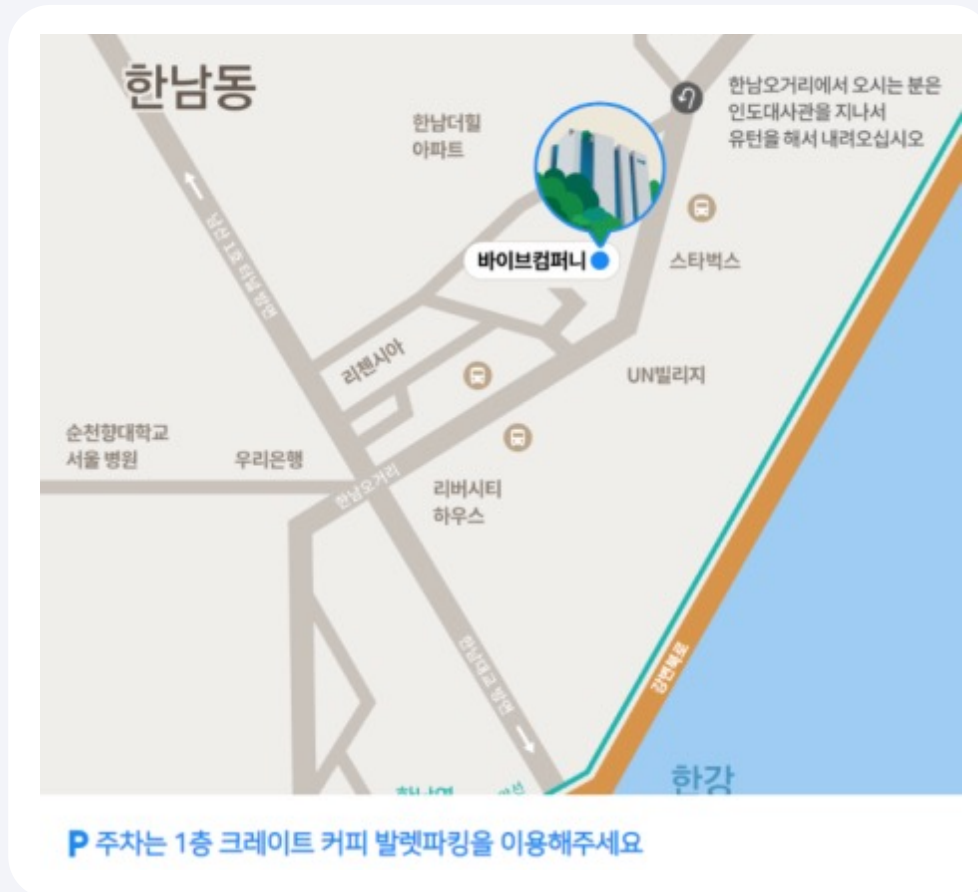
데이터 상품은 API를 통해 실시간 분석 결과를 제공받는 방식과 고객의 요구사항에 따른 데이터를 일 배치를 통해 파일 형식으로 제공하는 방식으로 구성



항목	바이브컴퍼니 썸트렌드 Social Data Service	바이브컴퍼니 썸트렌드 On Demand Data Service
제공 소셜 빅데이터	 트위터  인스타그램  블로그	 뉴스  커뮤니티/카페  유튜브
분석 키워드 수	키워드 제약 없이 무제한	300개 문서필터링 조건(포함어, 제외어)지원
제공 방식	HTTP, HTTPS 통한 API 요청을 실시간으로 처리	HTTP, HTTPS, SFTP 일배치 압축 파일 제공
제공 정보	주제 키워드의 문서수, 연관어, 감성어, 원문 정보 등	약정 키워드의 원문정보와 통계 정보 (문서수, 연관어, 감성어 등) 요구사항에 따라 추가 정보 제공
결과 형식	JSON	JSONL, CVS, TSV 등 협의에 따라 결정

05. 서비스 가격 정책

항목	바이브컴퍼니 썸트렌드 Social Data Service	바이브컴퍼니 썸트렌드 On Demand Data Service	비고
월 서비스 구독료	1개 채널 구독 시 400만원	1개 채널 구독 시 300만원	2개 상품 동시 구독 시 5개 채널 이상 1,500만원/월
	채널 추가 시 1개당 200만원 추가	채널 추가 시 1개당 100만원 추가	
	5개 채널 이상 구독 시 1,200만원	5개 채널 이상 구독 시 700만원	



Contact us

바이브컴퍼니 썸트렌드본부 | 세일즈팀

Address

서울시 용산구 독서당로97 바이브한남

E - Mail

sales@some.co.kr

Tel.

02-565-0531(내선 2번)

End of Document