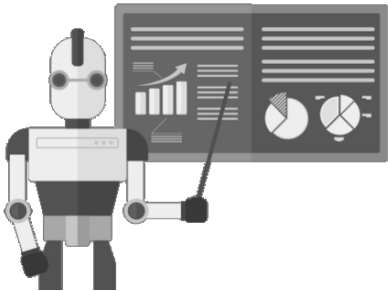


머신러닝 기반의 학습분석기술 적용 사례



기술, 교육의 거품을 걷다!

1 온라인 교육강의 변화 필요성

학습방법의 변화

2016년 전국 43개 고등학교 2,577명의 학생을 대상으로 실시한 'EBS 수능강의 성과분석' 연구

- 33.5%가 자기주도 학습 및 학교수업에 충실하여 EBS수능강의 시청이 불필요
- 응답학생의 29.3%가 더 좋은 사교육 때문에 불필요하다고 답변

구분	빈도	비율
EBS 수능강의 서비스가 우수하지 않아서	35	3.4%
혼자 공부하는 것이 효율적이어서	148	14.4%
학교 수업만으로 충분해서	196	19.1%
학원, 과외 등에서 EBS 수능강의 및 교재를 요약 정리해 주기 때문에	79	7.7%
사교육 때문에 시간이 없어서	222	21.6%
이용 여건이 안 되어서	39	3.8%
공부 자체에 흥미가 없어서	201	19.6%
기타	107	10.4%
합계	1,027	

<EBS수능강의를 이용하지 않는 이유에 대한 설문 결과>
출처 : EBS 수능강의 중단연구 보고서. 2016년 12월. 성신여자대학교 강태훈 등

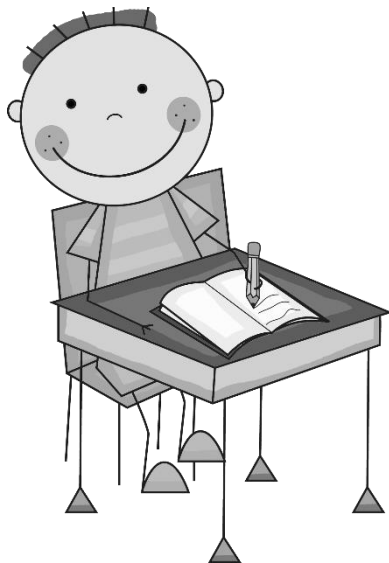
2 콘텐츠 자체의 변화필요

자기주도학습이 가능한 콘텐츠와 서비스로의 변화 필요

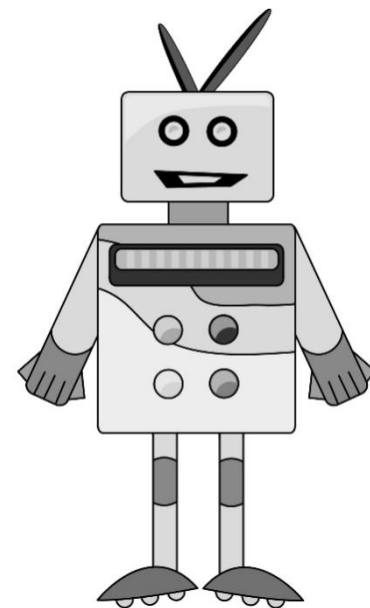
자기주도 학습

맞춤형 서비스

교육비용 저감

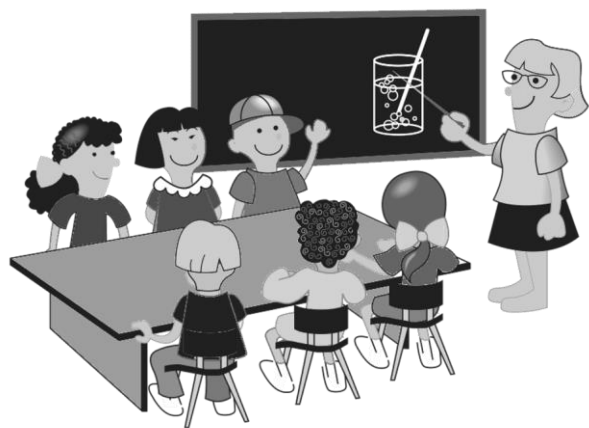


온라인 강좌 청취
강좌와 교재만으로
학습진행이 가능하도록 !

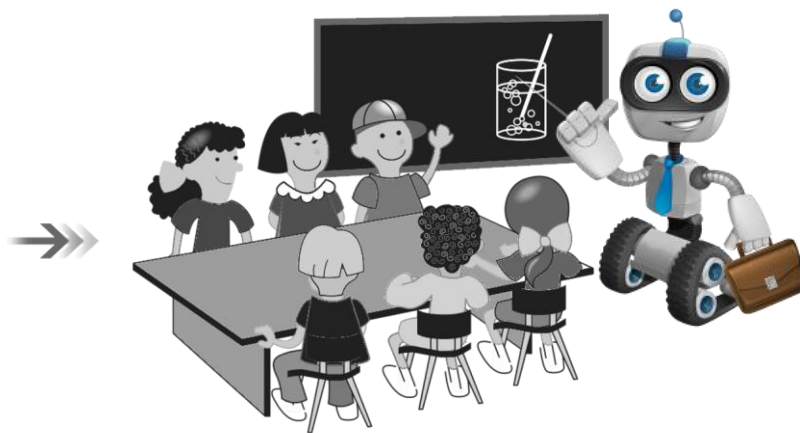


3 교육주체의 역할 변화

교사의 한계와 앞으로의 역할



학생 개개인에 대한 파악 어려움...



지능정보기술이 획일적인 수업방식 개선!

학생들이 스스로 학습하도록 유도

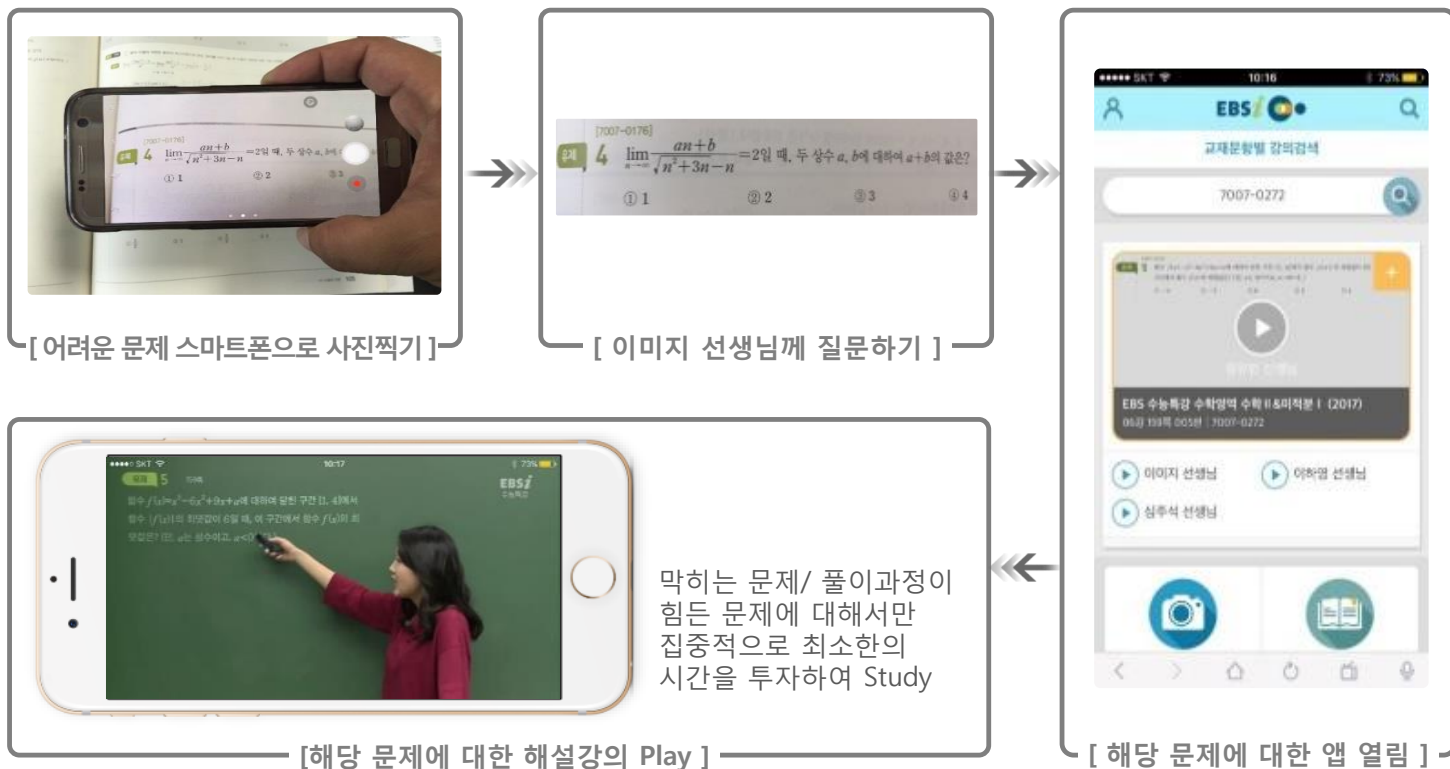
수행과정 및 결과에 따라 교사가 맞춤형으로 코칭...

4 EBS 지능정보기술 적용사례 #1

EBS 스마트북 검색서비스

- 이미지 검색기반 동영상 문제풀이 서비스
- 95% 이상의 정확도 확보

EBS 스마트북 서비스



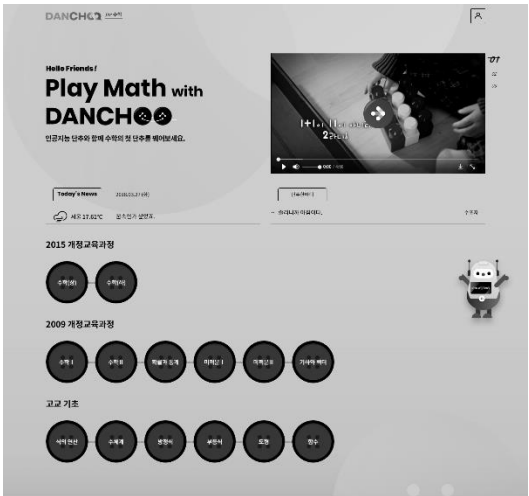
4 EBS 지능정보기술 적용사례 #2

EBS 인공지능기반 1:1 튜터링 시스템 (단추)

■ 챗봇 기반의 수학과목 1:1 튜터링 서비스



인공지능
1:1 맞춤형 학습서비스



EBS 단추 전용 챗봇 “ChooBot”



4 EBS 지능정보기술 적용사례 #2


EBS 인공지능기반 1:1 튜터링 시스템 (단추)

■ 학습자 개인별 학습분석 서비스

DANCHOQ for 수학

단추 포인트 354P 003


단추의 학습 분석



단추의 한마디

“ 개념학습 후 문제풀이 단계에서 자신감을 가지세요. 실전문제 전 단계를 충분히 학습해 주세요. 단추 학습에 대한 집중이 부족한 상태입니다. ”

단추 지수



나의 학습 패턴

크기에 따라 학습량을 나타냅니다. ● 노력이 필요해요. ● 조금 더 노력해 주세요. ● 잘하고 있어요.

수학(상)

수학(하)

다항식의 덧셈과 뺄셈 > 개념 학습 중

다항식의 곱셈 > 개념 학습 중

개념이해력 레벨 UP 하기

7

4 EBS 지능정보기술 적용사례 #3

인공지능 기반 맞춤형 학습서비스



1 프로젝트 목표

변화된 학습환경에서의 공교육 서비스 목표 설정

“ 사교육비 경감과 평등한 교육 기회 제공 ”

자기주도학습 지원 강화



개인 수준별
맞춤 학습



역량 기반
학습



자기주도 학습
한계 극복

2 서비스 개발범위

EBS 수능포털에서의 사용자 학습분석을 통한 개인맞춤형 서비스



사교육비 절감과

효과적인 자기주도학습 방법론 제시를 위해,

핵심 서비스 채택

- EBS 단추 학습분석엔진 전 교과 확대 적용
- AI 강좌추천 서비스
- 적응형 문제풀이 서비스
- 모바일 스마트북 수준별 문제풀이 서비스
- 개인별 학습현황분석 DashBoard
- 맞춤형 학습 플레이어

수능 연계 교재·강좌 운영, 연간 200만명, 일평균 약 28만명 이용

연간 이용자수

연도	이용자수
2017	2,087,265
2018	1,485,666

일평균 방문자수

연도	방문자수
2017	275,075
2018	278,830

운영 강좌수

연도	강좌수
2018	8,875

1인당 평균 수강 강좌수

연도	수강신청수
2017	8.8
2018	6.4

※ 2018년은 6월 기준

강좌·교재 검색, 동영상 학습, 문제풀이가 핵심 학습 서비스



핵심 서비스



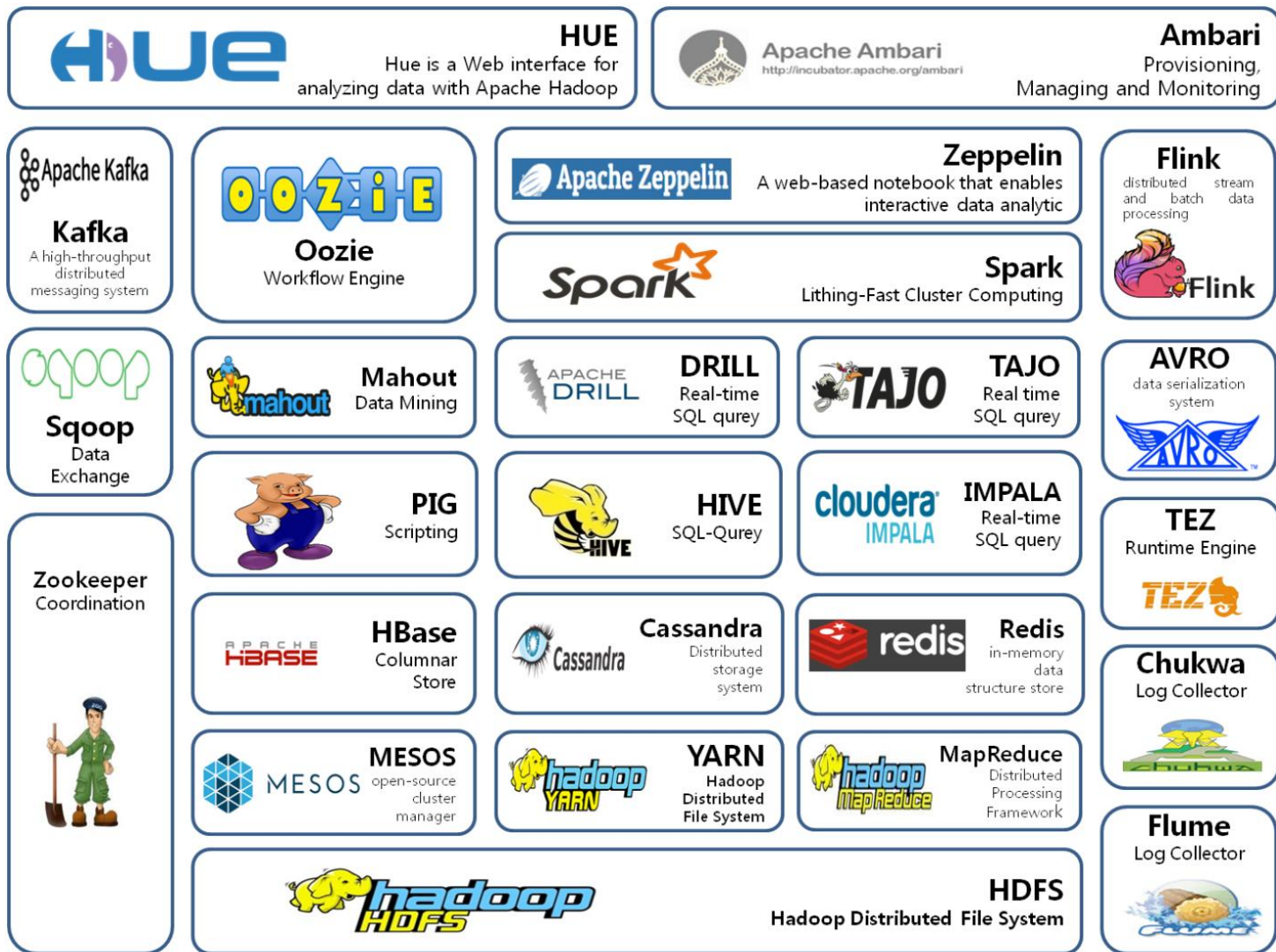
강좌/교재 검색

동영상 학습

문제풀이

4 학습분석 아키텍처

기존 빅데이터 분석체계



출처 : <http://1004jonghee.tistory.com>

01 실시간 데이터 처리 한계

- 하둡 MapReduce는 일괄처리(Batch) 방식이기 때문에 실시간 데이터 처리, 조회가 안됨

02 학습자 경험데이터는 실시간으로 누적, 분석되어야 함

- 누적 200만명이 넘는 사용자의 경험데이터가 실시간으로 쌓이고, 분석되어야 함

03 분석된 경험데이터는 추천의 형태로 서비스화 되어야 함

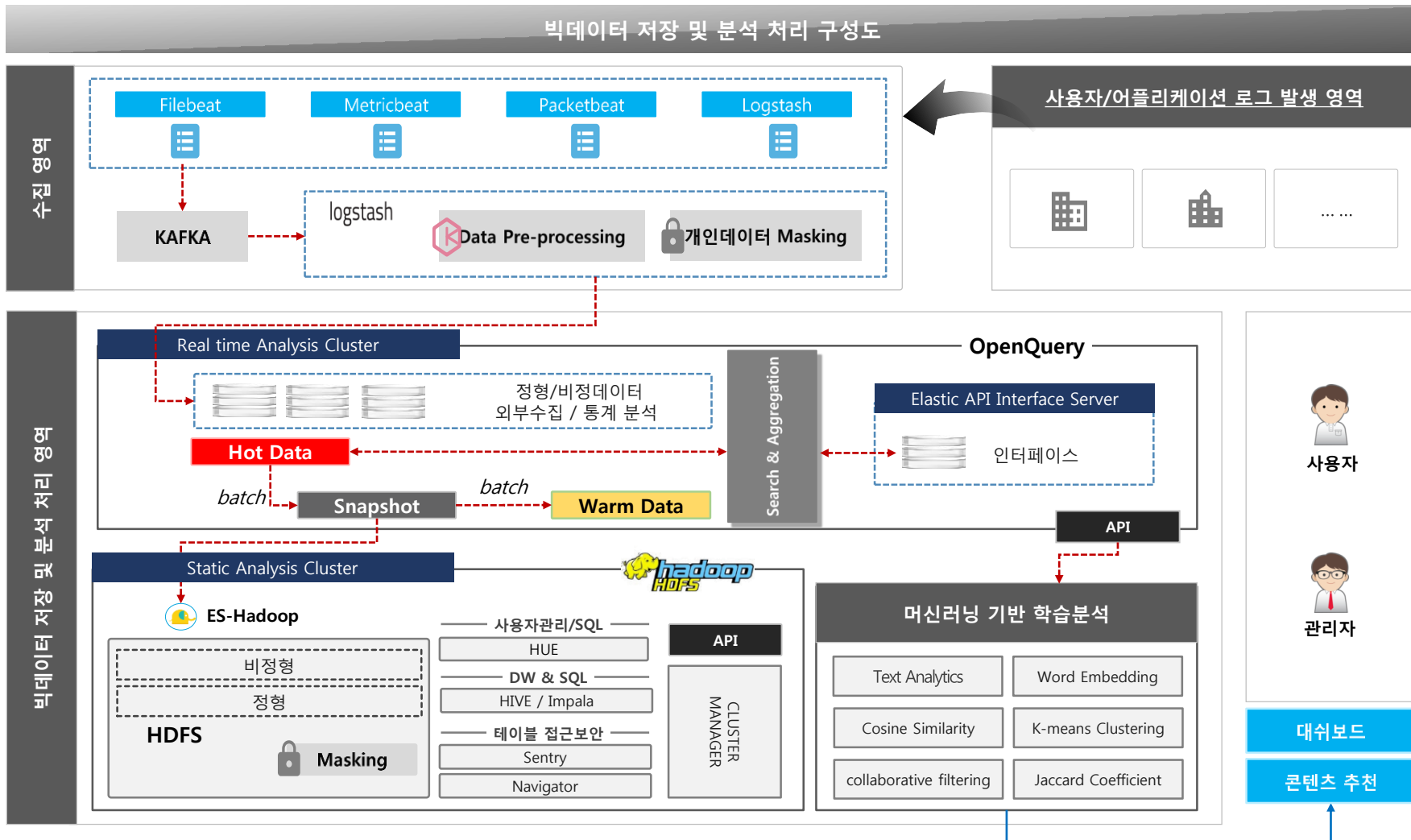
- 매일매일 누적된 경험데이터를 기반으로 분석이 이루어지고, 분석을 토대로 사용자 맞춤형 콘텐츠가 매핑

04 요구되는 분석을 위해 너무 많은 IT 전문가가 필요

- 하둡 Eco 환경에서 Hadoop, Hive, R, Spark, Yarn 등 EBS 가 필요한 분석을 위해 너무 다양한 전문가 필요

4 학습분석 아키텍처

실시간 빅데이터 분석체계

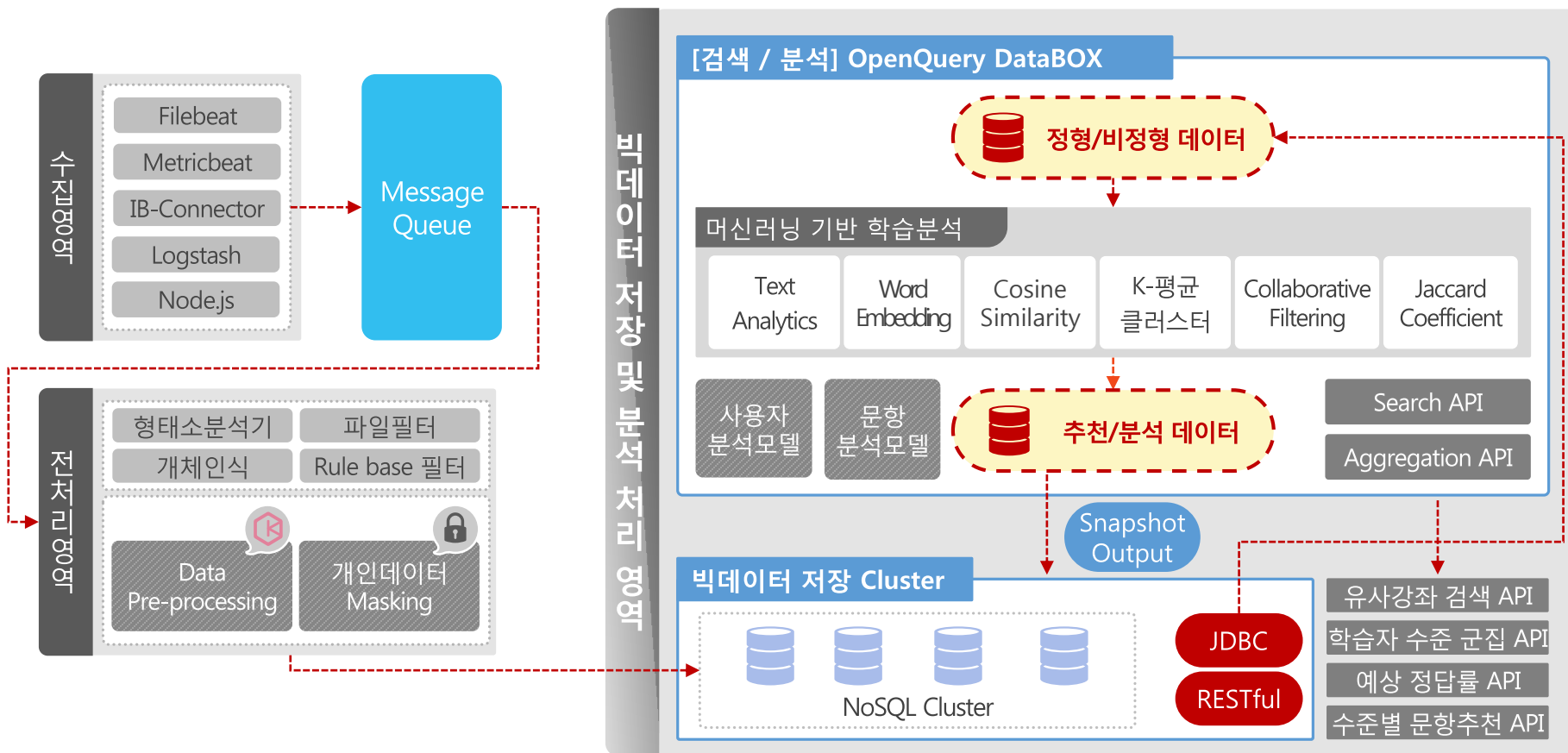


출처 : <http://1004jonghee.tistory.com>

4 학습분석 아키텍처

EBS 고교 전과목 빅데이터 분석체계 구축

정형/비정형 빅데이터 수집, 저장, 관리, 분석 체계 구축



5 사용자 경험데이터 수집

수능포털 실 환경에서의 사용자 경험데이터 식별

실환경에서는 로깅할 수 있는 사용자 경험데이터가 많지 않다!

분석 지표	정의	행위	분석 대상 Data	분석 알고리즘
회원별 관심도	<ul style="list-style-type: none"> 사용자 행위로 부터 분석된 강좌에 대한 회원별 관심도 <ul style="list-style-type: none"> - 상품 구매 전단계의 관심 행동 	<ul style="list-style-type: none"> 강좌 정보 조회 선생님 정보 조회 교재 정보 조회 맞보기, 수강이력 등 조회 	<ul style="list-style-type: none"> 유입경로 맞보기 조회수 강의 목록 페이지 조회수 수강후기 페이지 조회수 강좌 상세페이지 조회수 교재 검색 이력 교재 상세 조회 이력 강사 검색 이력 강사 상세 조회 이력 학습자 위치 정보 	<ul style="list-style-type: none"> 소비이력기반 협업 필터링
회원별 선호도	<ul style="list-style-type: none"> 구체적인 구매행위(수강신청)가 발생하는 단계로 실제 구매가 일어날 확률을 예측 	<ul style="list-style-type: none"> 강좌 수강 신청 동일 선생님 강좌 수강 	<ul style="list-style-type: none"> 강의별 수강정보 강좌 다운로드 Index별 수강 이력²⁾ 학습 Q&A 작성 이력 수강후기 작성 이력 선생님 홈페이지 방문 이력 선생님 프로필 조회 이력 	<ul style="list-style-type: none"> 정형 데이터에 대한 클러스터링 콘텐츠 소비이력 데이터
회원별 만족도 / 성취도	<ul style="list-style-type: none"> 제품의 활용 단계에서 발생하는 심리적 만족도 및 성취도 향상의 기여 정도를 예측 	<ul style="list-style-type: none"> 강의 수강 / 완강 학습 Q&A 작성 수강후기 작성 	<ul style="list-style-type: none"> 학습 Q&A 정량 분석 수강후기 정량 분석 강좌/강의 완강율 수강후기 비정형 분석 문항 풀이 결과 정/오답 문제풀이 소요시간 	<ul style="list-style-type: none"> 정량 데이터 분석 비정형 데이터에 대한 핵심 키워드 추출 구문 분석을 통한 긍/부정 평가
강좌 인기도	<ul style="list-style-type: none"> 시점별/기간별 강좌 수강율 변화 수강율 대비 완강율 	<ul style="list-style-type: none"> 수강 신청 완강 	<ul style="list-style-type: none"> 수강율 완강율 정답율 	<ul style="list-style-type: none"> 정량 데이터 시계열 분석

5 사용자 경험데이터 수집

수능포털 실 환경에서의 사용자 경험데이터 식별

수집경로	구분	수집 대상 Data	데이터 유형
행위이력 수신	강좌 조회	유입경로	이벤트 수집
		맞보기 조회	
		강의 목록조회	
		수강후기 조회	
		강좌 상세페이지 조회 횟수	
	교재 조회	학습자 위치 정보	
		교재 검색 이력	
		교재 상세 조회 이력	
	강좌 수강	강의별 수강정보	
		강좌 다운로드	
		Index별 수강 이력 ²⁾	
		학습 Q&A 작성 이력	
		수강후기 작성 이력	
	선생님 조회	선생님 홈페이지 방문 이력	
		선생님 프로필 조회 이력	
EBS고교 데이터	강좌 특성	강좌 수강율	정형(수치)
		학습 Q&A 정량 분석	비정형
		수강후기 정량 분석	
		강좌/강의 완강율	
		수강후기 비정형 분석	
		강좌 범위 / 특징	

수집경로	구분	수집 대상 Data	데이터 유형
행위 이력 수신	문항 조회	스마트북	이벤트 수집
		문항코드검색	
		강의 학습 이력	
	해설강의 시청	해설강의 학습 시간	
		문제 풀이 결과	
	문항 풀이 결과	정/오답	이벤트 수집
		문제 풀이 소요시간	
		교과분류	
	EBS고교 데이터	태그	정형
		지문	비정형
	문항 특성	문제	정형(수치)
		난이도	
		정답율	
		보기선택 비율	

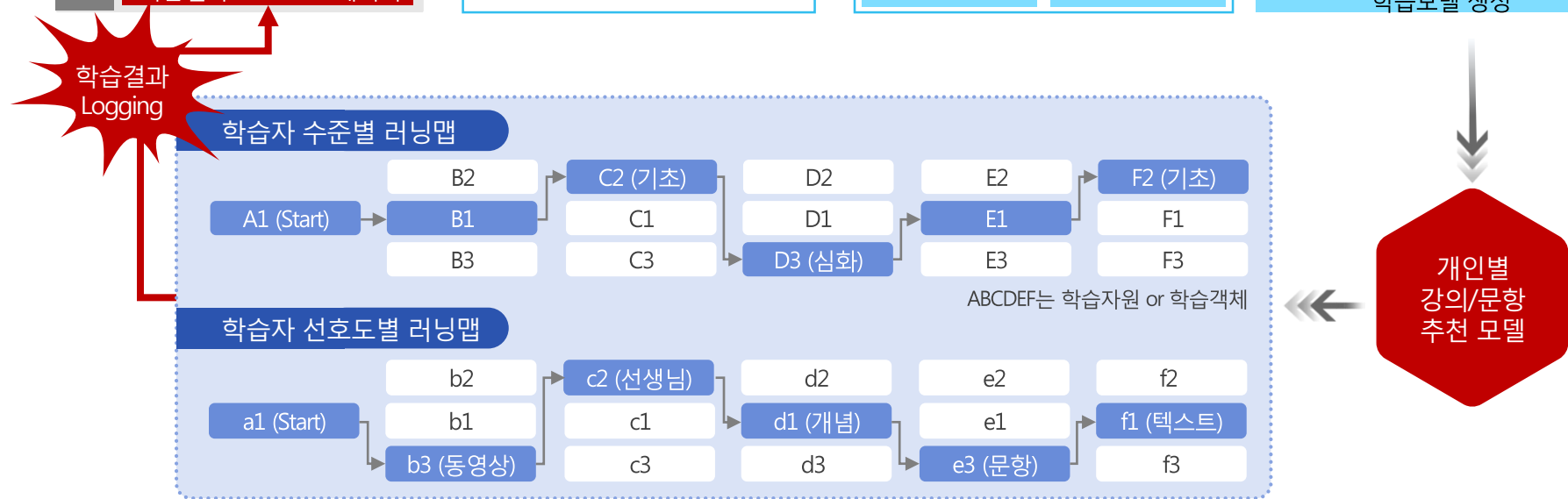
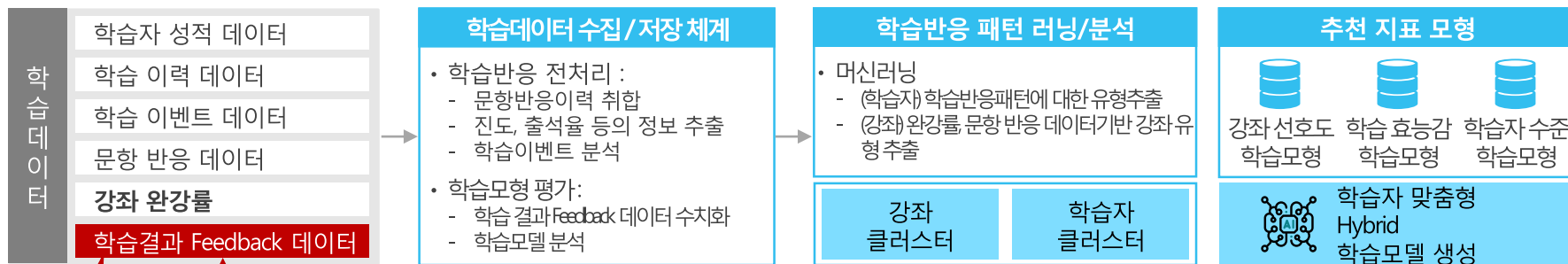
경험데이터 수집목적? 분석툴?

- 학습자 과목별 수준분석 및 관심도 분석?
- 학습 성취도에 영향을 미치는 학습패턴?

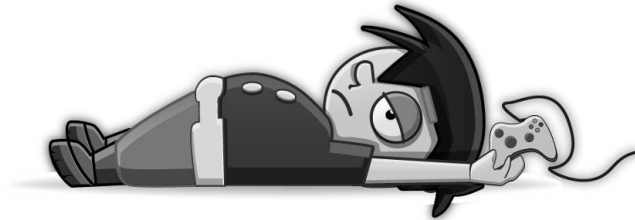
6 학습분석 및 추천 서비스

머신러닝 기반의 맞춤형 추천서비스

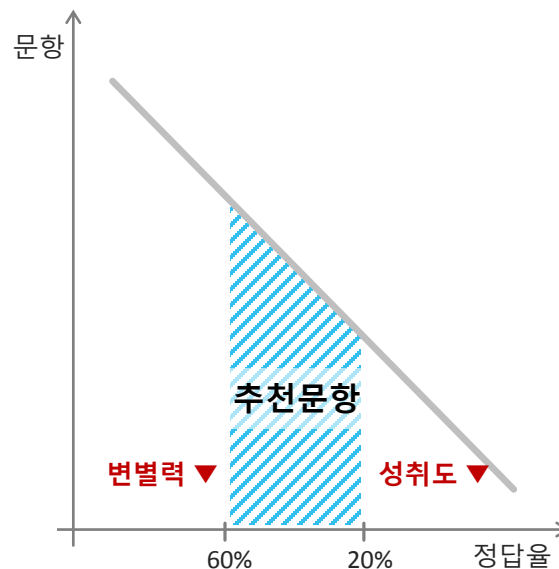
학습경험데이터를 활용한 머신러닝 기반 추천모형 개발



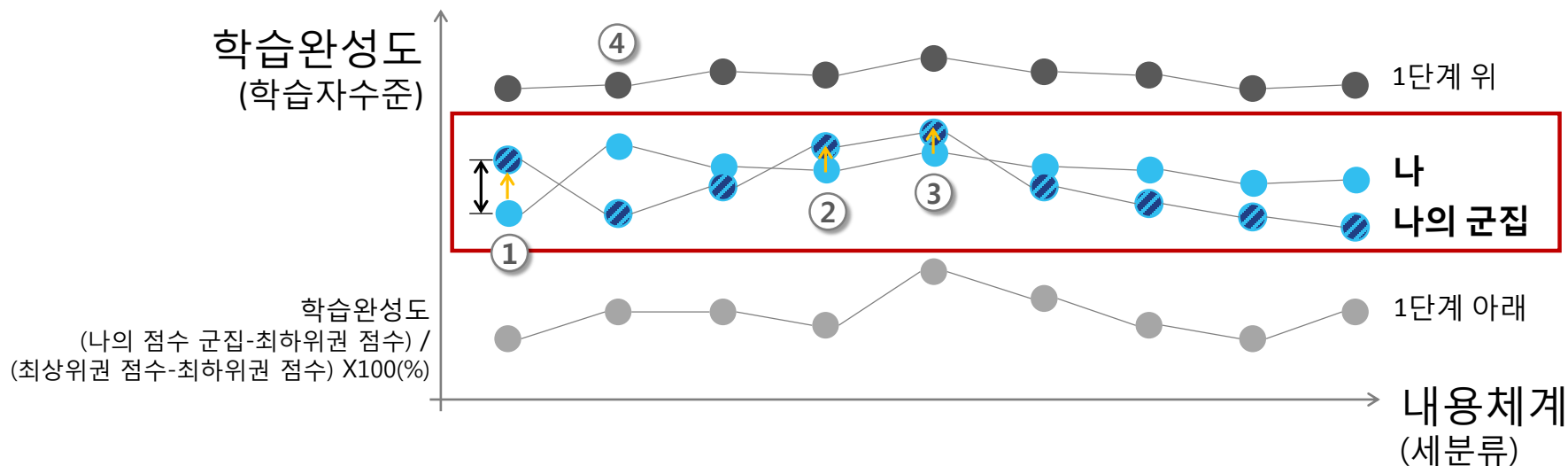
아무리 좋은 기능의 서비스를 만들어도
학습자가 효과가 없으면 그만!



동일 그룹 학습자 분석을 통한 선별적 문항 추천



성적대별 사용자 그룹을 10grade로 구분, 학습자 그룹의 상하계적에 따른 콘텐츠 추천



1) 학습완성도 평가 및 목표 지정 : 학습완성도 평균 분석을 통해 2가지를 판단

* Level 상승 : 일부 단원의 취약영역 보충을 통해 1단계 상승이 가능한 학생

* Level 보충 : 일부 취약영역을 보완하여 상승을 준비하는 학생

2) 학습 목표에 따라 문항 GAP분석 결과에 따른 문제를 추천

* 학습자가 속한 군집 기준과 편차를 계산하여 취약한 세분류의 문항 학습

* 모든 세분류에 대해 군집 기준을 상회 할 경우 1단계 향상 학습 수행

7 추천서비스 모형

학습자 문항반응 패턴별 예측 모형

학습자별 문항반응에 따른 정/오답 예측



성취도 및 학습패턴의 유사성을 가지는 사용자 군집



유사패턴의 사용자 군집에서 문항별 정/오답 패턴분석



군집A



군집B



군집C

성취도에 영향을 미치는 학습패턴

콘텐츠 클러스터

연관 규칙 학습
(FP-Growth 알고리즘 사용)

User2와 User5의 경우
학습패턴도 유사한데,
문항반응 또한 유사

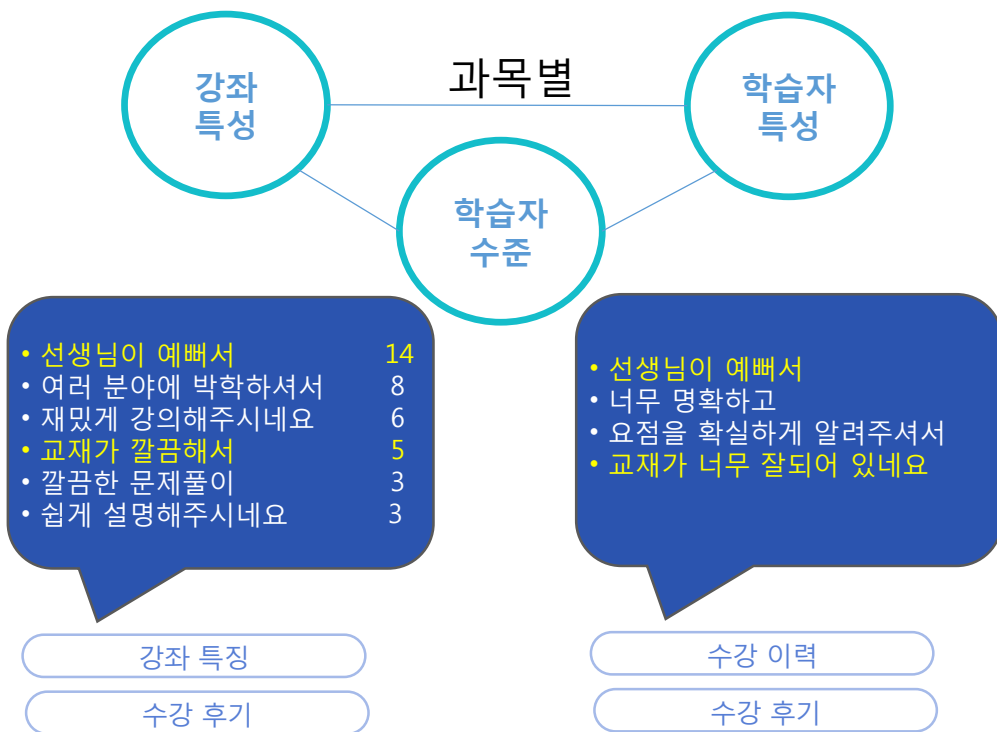
User5의 경우 문항3과
문항9가 포함된 학습주제에
취약할 가능성이 매우 높음

	문항 1	문항 2	문항 3	문항 4	문항 5	문항 6	문항 7	문항 8	문항 9
User1	T	T			T			F	
User2	F	F	F	T		F	T		F
User3	T		T		T		T		
User4	T						T		T
User5	F	F		T		F	T		
User6		F	T	T			T		
User7		T	F	F	F	T		T	

< 문항풀이 이력 >

Discourse Analytics 를 활용한 학습자-강좌 추천

학습자의 수준과 선호도를 고려한 강좌추천



강좌

- 수강이력, 강좌특징, 수강후기, 학습 Q&A 등 비정형 데이터에 대한 사전 기반의 주제어 추출과 담화분석을 통한 긍/부정을 추출하여 유사 특성 강좌 클러스터링을 수행

학습자-특성

- 학습자의 수강 패턴을 로깅하여 수강 참여도, 집중도, 관심도를 수치화

협업필터링을 활용한 선호도 기반 강좌 추천 모형

사용자-강좌 선호도 연산을 위한 매트릭스

강좌 ID	미적분Ⅱ 이만형	심주석의 미적분	차현우의 수학가형	정승제의 확률과통계	차현우의 기하벡터	최은진의 미적분Ⅰ
김태희	1	0	0	0	0	1
장동건	0	1	1	0	0	0
박나래	0	1	0	1	1	1
이만호	0	1	0	1	1	0
김희선	1	1	0	1	0	0
박보검	0	1	1	0	1	1

로그인한 학습자 학습이력

0	0	1	0	0	1	1
---	---	---	---	---	---	---

Description

- 학습이력을 바탕으로 학습한 강좌는 1, 학습하지 않은 강좌는 0으로 표현하여 좌측과 같은 matrix 구성
- 학습자 ID 중심으로 하는 row는 특정 학습자의 학습이력을 나타냄
- CF 방법론에서 널리 사용되는 Jaccard Coefficient(자카드 계수)를 사용하여 학습이력이 유사한 학습자들을 군집

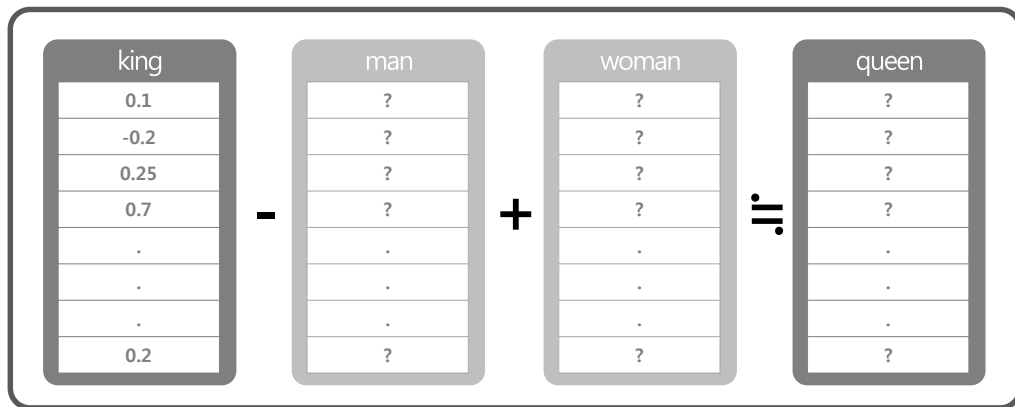


$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

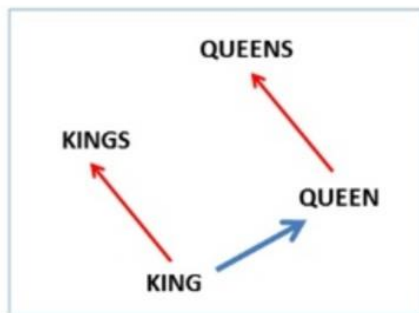
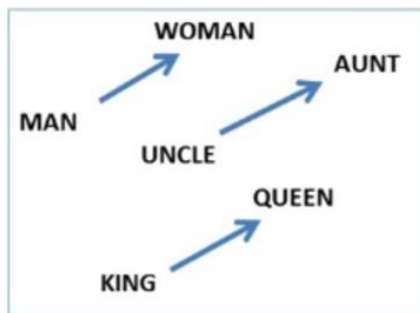
Word Embedding이란

문서집합에 존재하는 모든 단어들의 연관성을 수식에 따라 연산하여 특정 단어를 N차원의 Vector 값으로 매핑시키는 것

Word2Vec의 의의



$$\text{vec}(\text{"king"}) - \text{vec}(\text{"man"}) + \text{vec}(\text{"woman"}) \approx \text{vec}(\text{"queen"})$$



Description

- Word embedding의 한 종류인 Word2vec은 2013년 구글의 Tomas Mikolov가 발표
- Word embedding은 그 이전부터 존재해온 개념이지만, word2vec 이후부터 실필드에 활발히 적용하고 있음
- 텍스트를 구성하고 있는 단어간의 관계가 파악된다면, 그 텍스트를 분석하는데 유리
- Word2vec의 결과물인 단어와 그 단어 vector값을 연산하면 좌측 상단과 같은 결과를 얻을 수 있음(2013, Mikolov)
- Word vector를 Vector space에 사상해보면 왼쪽 아래의 그림과 같이 연관된 단어는 벡터가 유사한 방향성을 갖고, 비슷한 속성끼리는 벡터의 차가 비슷한 성향을 나타낸다고 구글에서 발표함

8 콘텐츠 유사도 분석모델

강좌 및 문항 콘텐츠 유사도 분석 – Word Embedding

Word Embedding이란

문서집합에 존재하는 모든 단어들의 연관성을 수식에 따라 연산하여 특정 단어를 N차원의 Vector값으로 매핑시키는 것

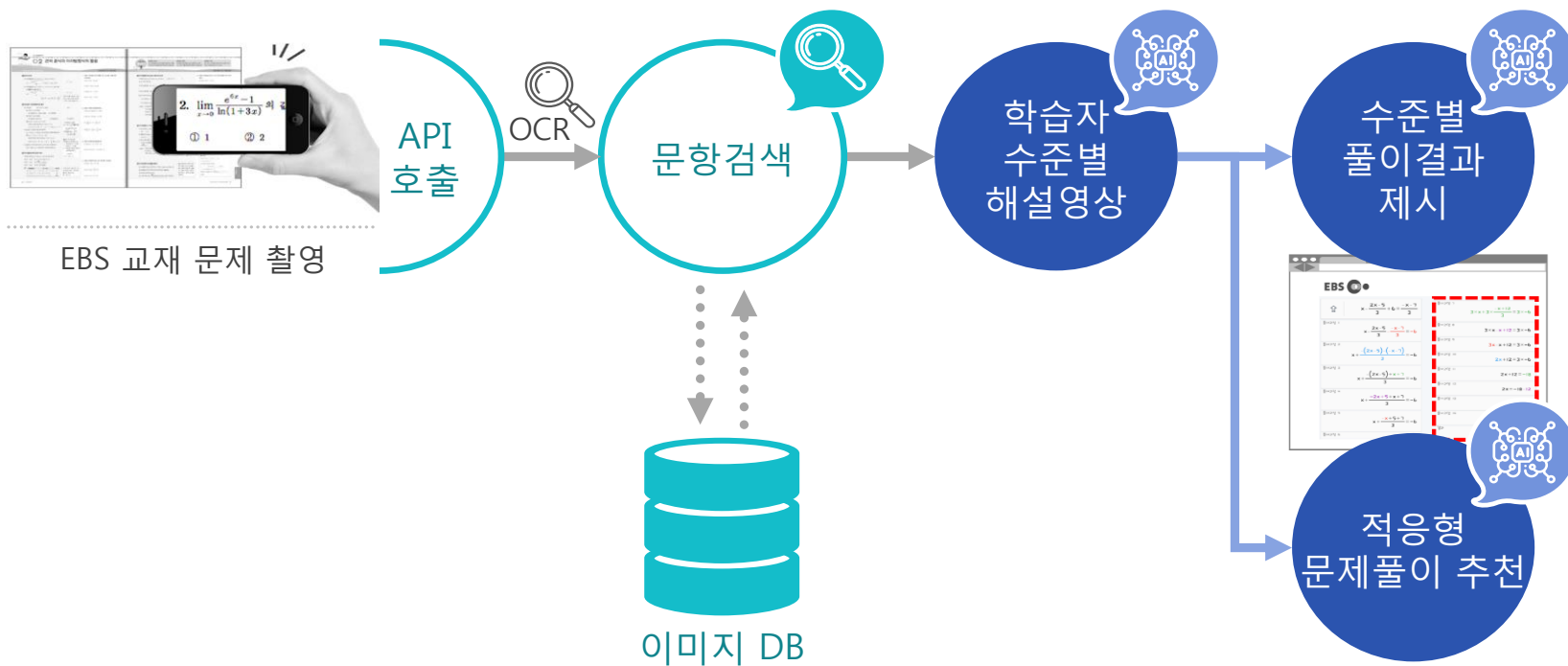
Word2Vec의 의의

문태 -0.6425338387489319 0.12770740687847137 -0.6745184659957886 -0.7260578274726868 -0.11634363234043121 -0.45477238297462463 0.81771081686019
수집필터 0.6624175906181335 -0.7728039026260376 0.3481932282447815 -1.0785865783691406 -0.7994920015335083 -1.119893193244934 1.5388962030410767
인식고취 -0.020183173939585686 0.2059366106986995 0.24498167634010315 -0.022216467186808586 -0.907741904258728 -0.6645400524139404 1.5123648643
자원 -0.49366915225982666 -0.5618546009063721 0.7772853970527649 0.14982593059539795 0.9098257422447205 -0.8767074346542358 0.48125022649765015
정형력 0.03822244703769684 -0.5345596671104431 -0.21268121898174286 -0.4816787838935852 -0.8646307587623596 0.34037888050079346 -0.131214186549
출제자의 0.13254088163375854 0.820777142524719 -0.10214882344007492 -1.2180293798446655 0.04459177702665329 0.37551799416542053 0.88250380754470
구대명달자 0.8223997354507446 0.13056111335754395 -0.3701082468032837 0.17201228439807892 -1.2619986534118652 0.11351416260004044 0.2134982049465
저장 1.000612497329712 -0.4331715703010559 0.13464726507663727 -0.18741054832935333 -0.5055986046791077 -0.4831930100917816 -0.5892481207847595
느낌전달 0.15751416981220245 -0.3400643765926361 -0.044380683451890945 0.7930171489715576 0.044797513633966446 0.008602166548371315 -0.065719641
숙목과 1.1580184698104858 0.985427975654602 0.26559242606163025 0.017504503950476646 -0.05304703861474991 -0.30301815271377563 0.15947166085243
보안관리실무과정 1.3442926406860352 0.6099341511726379 0.35259151458740234 0.49244239926338196 0.3359580934047699 0.1862952411747742 0.38766479492
회사원들 -0.5857143402099609 0.9533464908599854 -0.255487322807312 -0.9016402363777161 0.5104916095733643 0.3215186595916748 0.23227834701538086
cad 0.22319425642490387 0.1615723967552185 0.024212459102272987 -0.9088289737701416 -0.28662601113319397 -0.19596953690052032 -0.427324235439
국제발표 0.017406238242983818 -0.336251437664032 -0.017428722232580185 -0.3752767741680145 0.8509261608123779 -0.7507850527763367 0.71560859680
전외 0.47082576155662537 -0.18272724747657776 0.5099952816963196 0.15185759961605072 0.3498046100139618 -0.8438832759857178 0.2723866105079651
직장인들 0.3219662308692932 -0.029867682605981827 -0.03558465093374252 -0.20510295033454895 -0.37495747208595276 -0.47543099522590637 0.36539539
실무교육 0.6238865852355957 0.634814977645874 -0.14051315188407898 -0.29194968938827515 -0.49324682354927063 -0.34403863549232483 0.371834456920
키즈 0.6810885071754456 -0.13626648485660553 -0.056554120033979416 -0.6479544043540955 0.9398791193962097 -0.5529394149780273 -0.89887309074401
전문비서 0.3152894079685211 -0.2688623070716858 0.7611725330352783 -0.5008858442306519 -0.3127983510494232 0.031100153923034668 0.49353790283203
drama 0.48412075638771057 0.6063968539237976 -0.3628762662410736 0.14066150784492493 0.09592422097921371 -0.5424417853355408 0.044916465878480
통승 0.10530410706996918 -0.14211434125900269 -1.029761791229248 0.5703765153884888 -0.0020467594731599092 0.303449809551239 1.2389681339263916
지속경영 0.5098698139190674 1.8492283821105957 -0.0024949845392256975 -0.5352218747138977 -0.012715114280581474 -1.3011833429336548 0.325943499
경향과 0.7129415273666382 -0.11486132442951202 -0.28856799006462097 -0.3491358160972595 -0.4632704257965088 -0.9840283393859863 0.11701307445764
변화수도 -0.519091010093689 0.16697131097316742 0.7204274535179138 -0.5564998984336853 0.7372622489929199 -0.44884738326072693 1.263365745544433
flex 0.6432610750198364 -0.3566683530807495 -0.7349429726600647 -1.0670777559280396 -0.100922591984272 0.4888564944267273 0.6317049264907837
생활영어 0.28711673617362976 0.43301597237586975 -0.4930979907512665 0.7463801503181458 0.46233507990837097 -0.9045056700706482 0.06760358810424
bld 0.6201609359739509 0.30965475933405955 0.07301394505509703 0.13200730708543453 0.405083008479736 0.13637324008370764 0.731333703390631
유료전력 -0.4378354251384735 0.25568902492523193 -1.6003299951553345 0.884989857673645 -0.21550796926021576 -0.7906783223152161 0.29404455423355
cbj 0.610850600011581 0.3755553065730400 0.1146253204315667 0.4136701731785005 0.740533336880038 0.01257656600444046 1.051736970305680
중심 0.610850600011581 0.3755553065730400 0.1146253204315667 0.4136701731785005 0.740533336880038 0.01257656600444046 1.051736970305680
전인 0.610850600011581 0.3755553065730400 0.1146253204315667 0.4136701731785005 0.740533336880038 0.01257656600444046 1.051736970305680
계획 0.610850600011581 0.3755553065730400 0.1146253204315667 0.4136701731785005 0.740533336880038 0.01257656600444046 1.051736970305680
cbo 0.7704006433486938 0.6605299711227417 -1.113968226242005 -0.3102042769950887 0.02312164498238564 0.12861014902591705 0.538037061691284

Description

- 단어와 그 단어의 수치 벡터 샘플
- Dimension(벡터의 크기) 100, iteration(인공신경망 반복학습 횟수) 100을 적용한 샘플
- 맨 왼쪽이 unique 단어이고 차례대로 단어 벡터값이 나열되어 있음

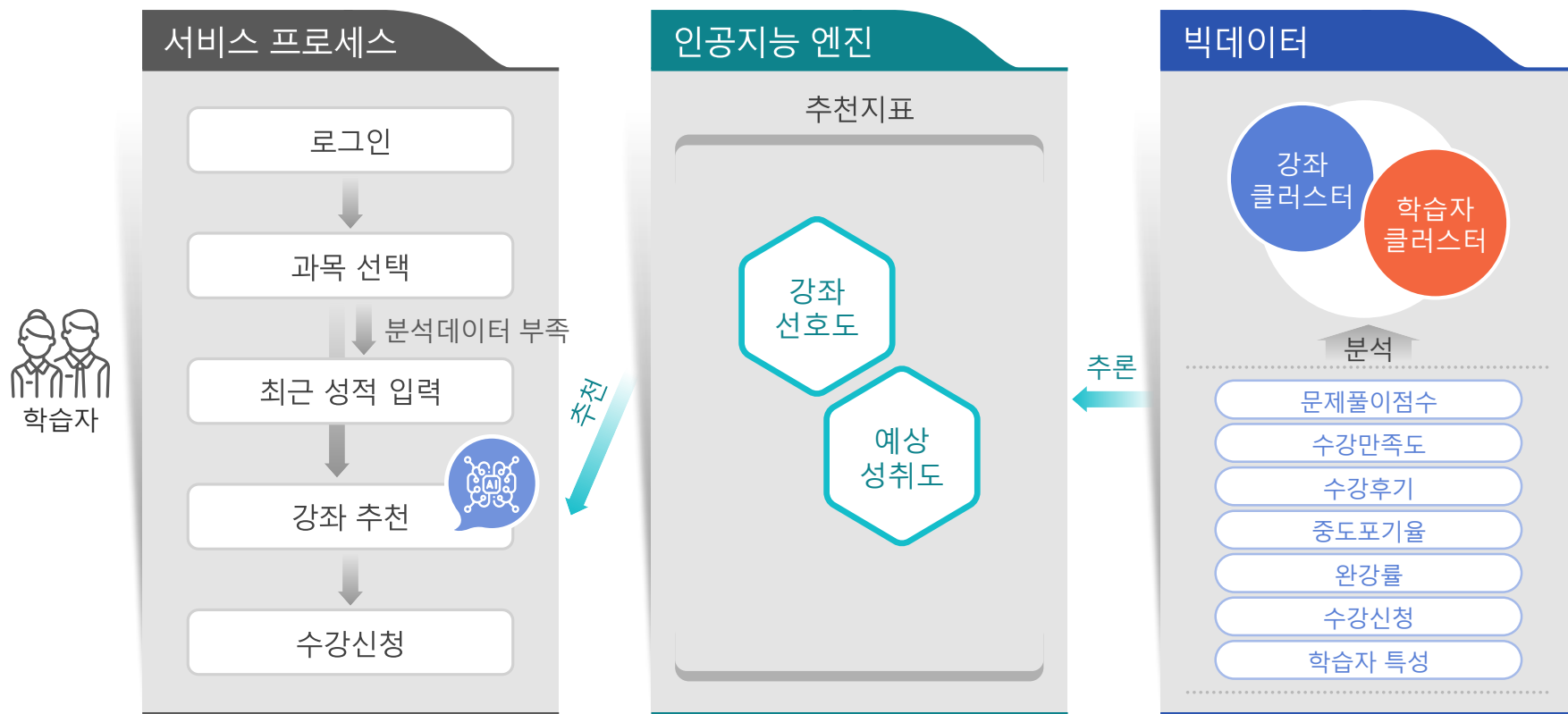
교재 촬영을 통한 EBS 모바일 수준별 문제풀이 서비스



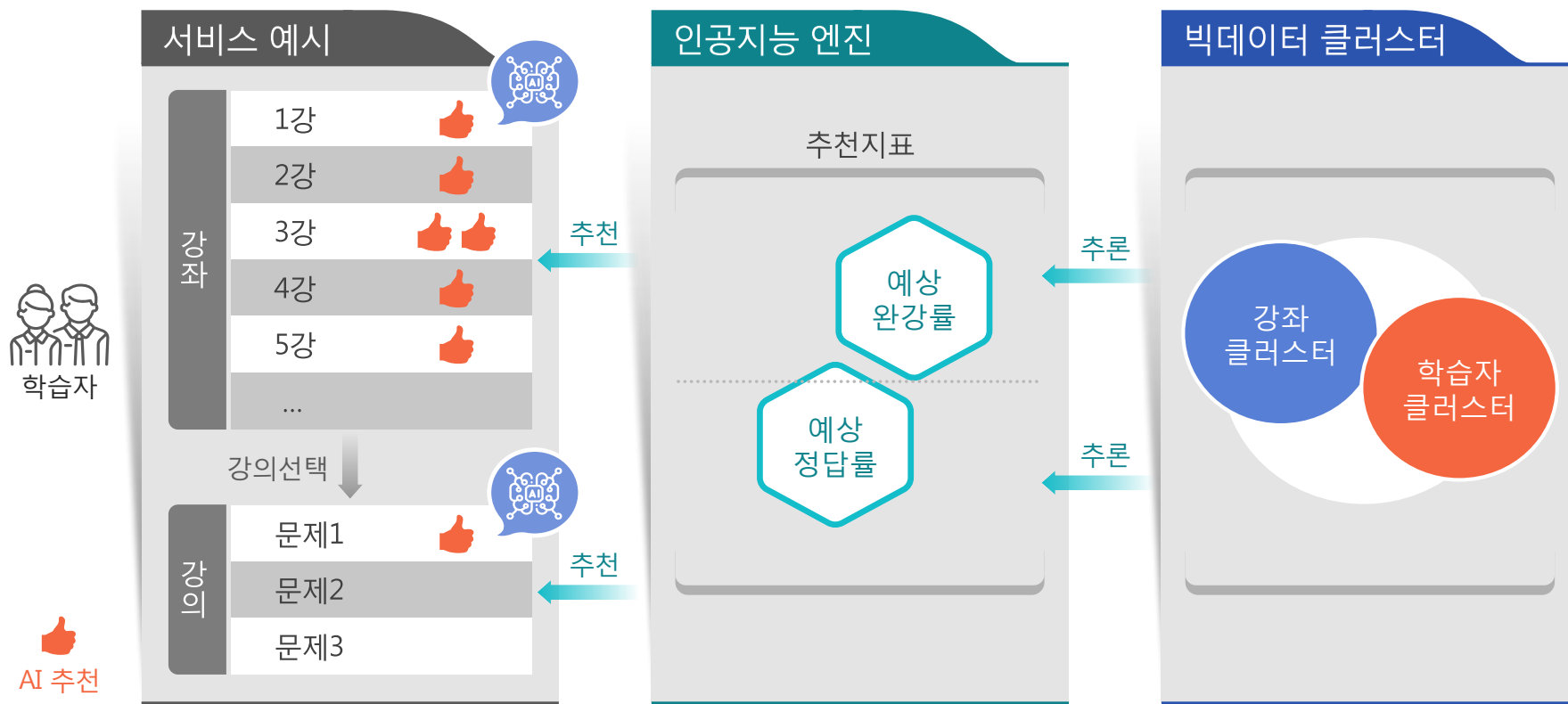
9 EBS 수능 실서비스 구현

AI 기반 강좌추천 서비스

학습자의 성취도, 선호도 등의 예측을 통한 맞춤형 강좌 추천 서비스



강의 및 문항별 완강률, 정답률 예측을 통한 맞춤형 학습추천 제공



학습자의 수준에 따라 문항의 난이도를 조절하여 적응형 문제풀이 제공



학습수준 분석도구

과목별/ 챕터별 학습수준 분석

추천

추천 강좌

해당 과목/챕터 우수성과자 (단기성과자)
의 러닝맵 조회

학평/모평 예측분석도구

학평/ 모평 취약문제 분석

추천

오답 예측에 따른 추천강좌

학습자 맞춤형 문제은행

10 학습분석의 전제조건

콘텐츠 구조에 종속적인 학습분석체계

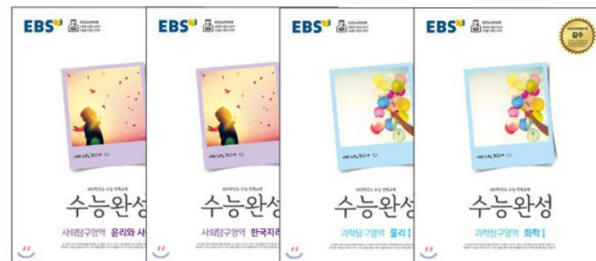
맞춤형 학습 알고리즘/ 솔루션?

콘텐츠와 운영데이터를 무시한 분석기법은 존재하지 않음



학습체계 개발, 알고리즘 적용은
자사 보유 콘텐츠에 종속되는 구조

“알파고에게 퀴즈 가르치면
퀴즈 챔피언이 될까?”



본 사업의 기반은 EBS 콘텐츠!

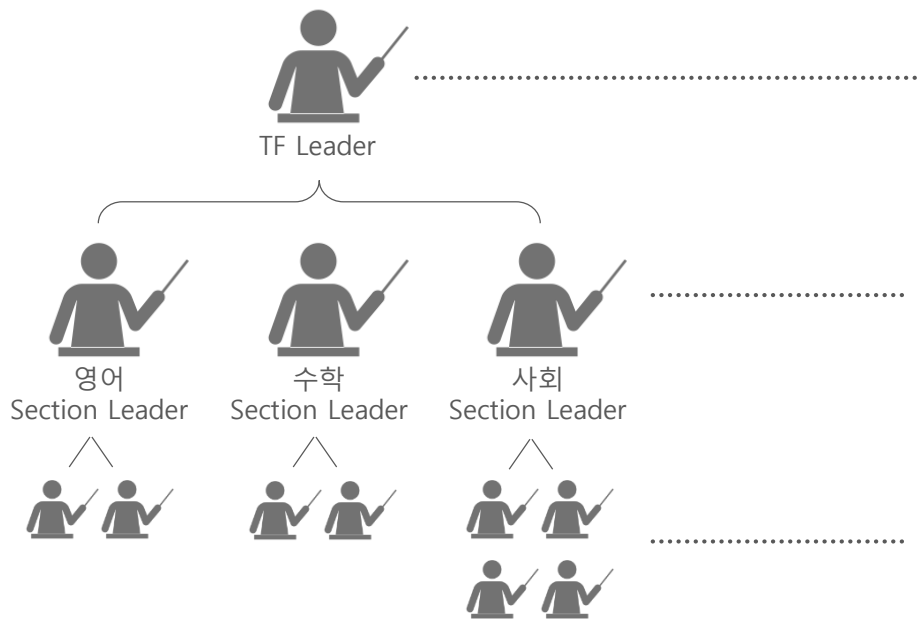
따라서, 학습분석체계는
EBS ● ● **교재 데이터에 종속적**



10 학습분석의 전제조건

전문교수자의 사업참여 필수

교육전문가 TF



- 과목별 학습체계도 개발
- 머신러닝 알고리즘 결과에 대한 실용성 검토
- 수능예측 및 예상 등급컷 설정 검토
- 맞춤형 학습체계 동작 검수
- 문항 메타데이터 카드 작성
- 개념학습 분절 및 메타데이터 카드 작성
- 문항 난이도 산정, 해설 난이도 산정

“실 교사 그룹의 안정적인 참여와 운영이 사업의 중요 포인트”



EBS 인공지능 기반 맞춤형 학습 서비스
**실질적인 자기주도학습에
도움이 되도록 하겠습니다.**