

# HyperCLOVA한국어 Fact 검증을 위한 자동 Claim 데이터 생성

*이종현<sup>01</sup>, 나승훈<sup>1</sup>, 신동욱<sup>2</sup>, 김선훈<sup>2</sup>, 강인호<sup>2</sup>*

*<sup>1</sup>전북대학교, <sup>2</sup>네이버*

1. 서론

2. 한국어 Fact 검증 데이터 생성

3. 실험

4. 결론

# 1. 서론

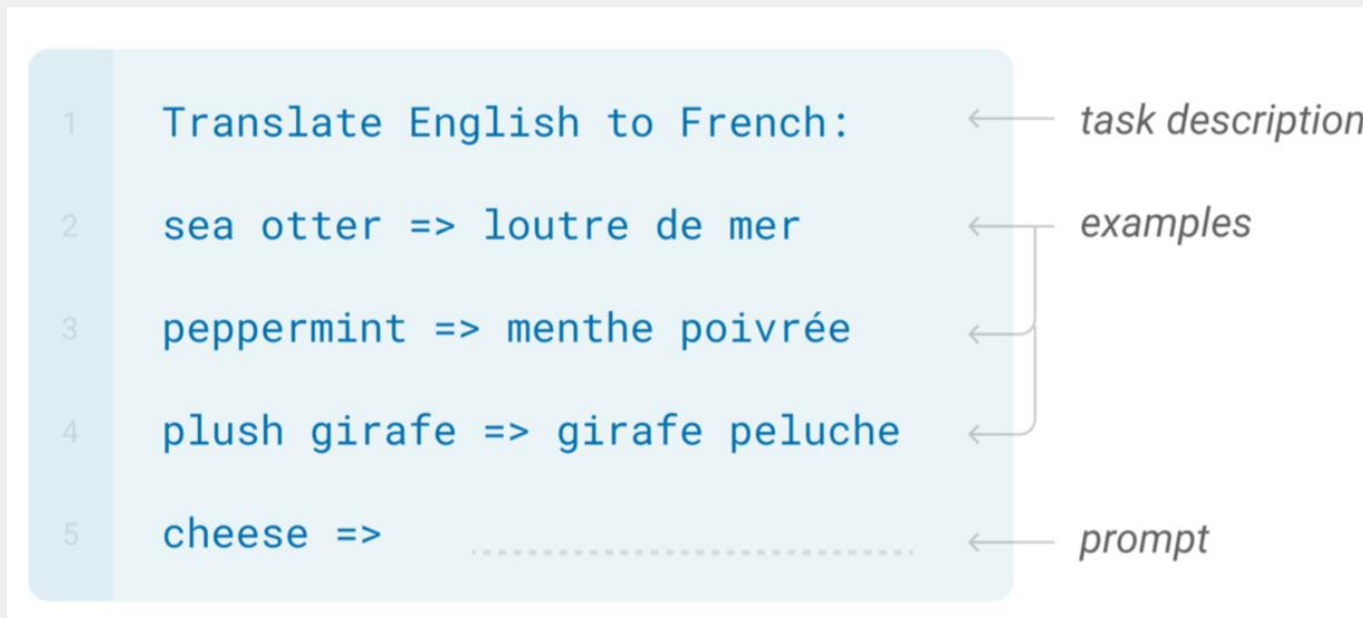
## ■ 사실 검증 (Fact Verification)

- 웹(Web) 상에서 쉽고 빠르게 수많은 정보를 얻을 수 있고 이러한 정보를 쉽게 타인과 공유할 수 있는 현대 사회에서 사실인 정보와 그렇지 않은 정보를 판단하는 것은 매우 중요함
- 국내에는 아직 자동 사실 확인 모델 구축을 위한 데이터 셋이 많이 부족한 상황임
- 사실 검증 절차
  - 1) 주장(claim)을 뒷받침 할 만한 증거 문서 검색
  - 2) 검색된 문서들 중 적절한 증거 문장 선택
  - 3) 뉴럴 모델을 통한 주장의 사실 확인

# 1. 서론

## ■ HyperCLOVA[1]

- 최근 매우 좋은 성능을 보이는 GPT-3 언어 모델을 한국어 태스크에 맞게 학습 시킨 모델
- 모델에 태스크의 설명과 few shot 을 넣어주어 예측 값을 얻어냄



출처: <https://ai.stanford.edu/blog/in-context-learning/>

## 2. Fact 검증을 위한 데이터 셋 구축

### ■ Fact 검증 데이터 셋 자동 구축

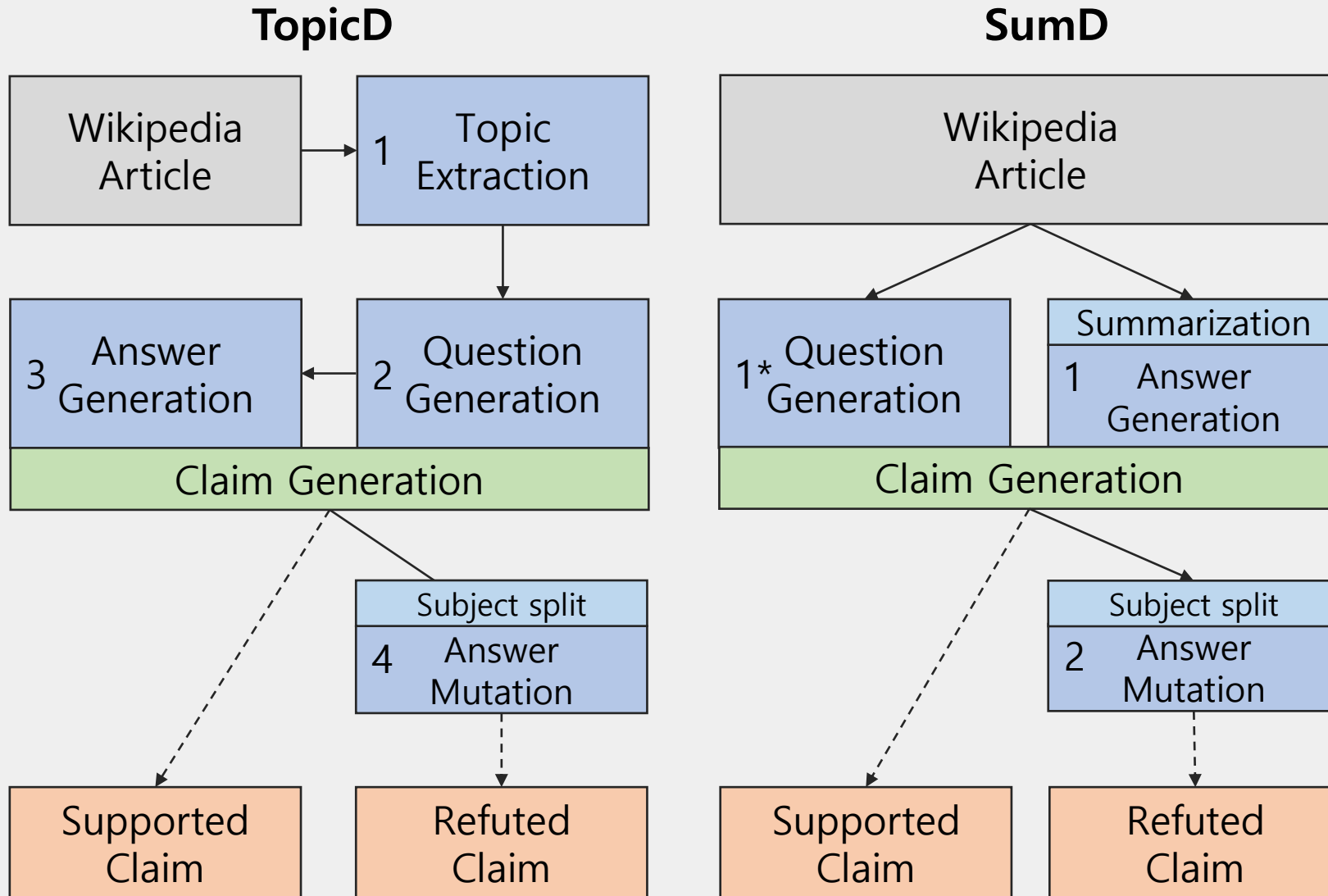
- Fact 검증 데이터 예시

Question	Answer
세계에서 국토가 가장 큰 나라는 어디야?	세계에서 국토가 가장 큰 나라는 러시아입니다. 총면적 1,709만km2로 세계에서 가장 큰 국토를 가진 나라입니다.
True Evidence	
러시아: 총면적 1,709만km2로 세계에서 가장 큰 국토를 가진 나라이며, 지구 육지 면적의 7분의 1을 차지한다. 국토가 넓어 러시아 안에서도 시차가 11시간까지 나며, ...	

- 사실 확인 데이터의 Label 은 SUPPORTED 혹은 REFUTED.

## 2. Fact 검증을 위한 데이터 셋 구축

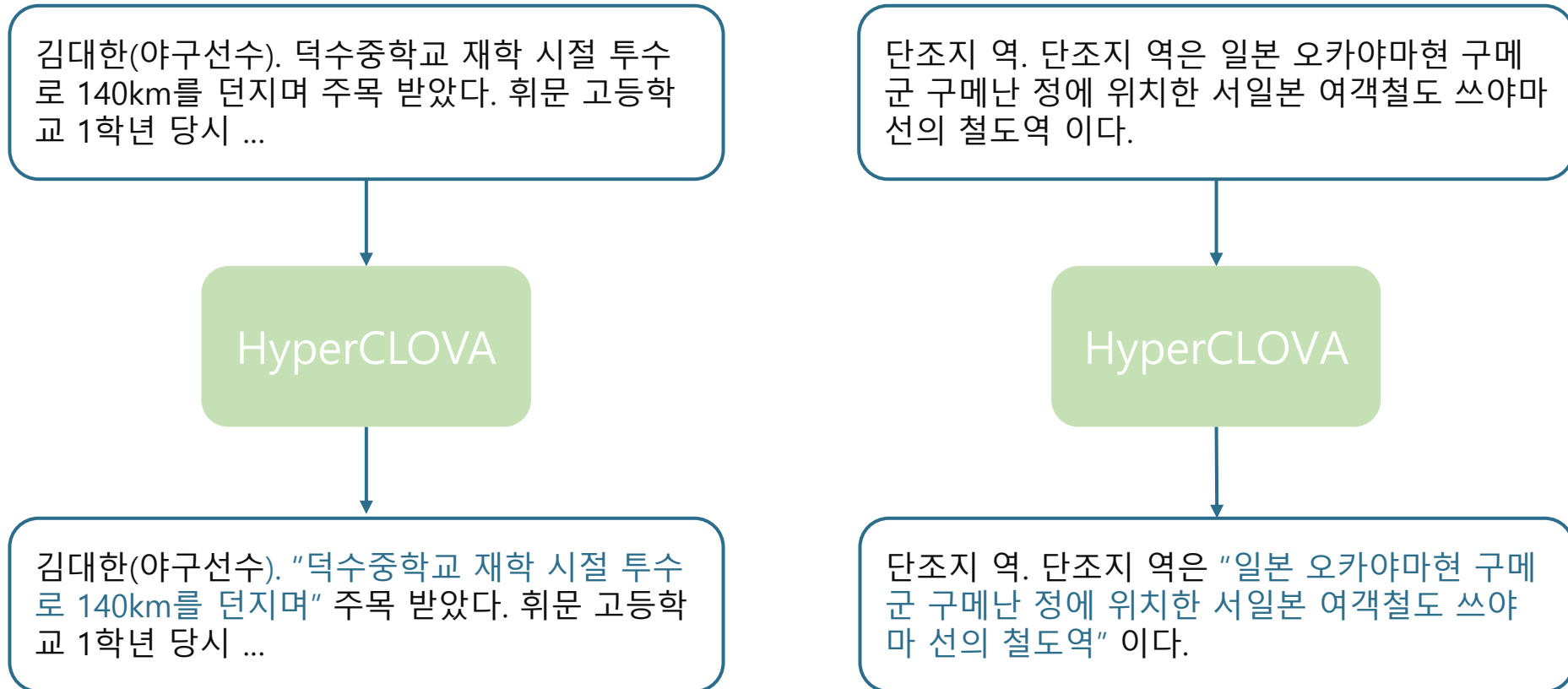
### Fact 검증 데이터 자동 생성 절차



## 2. Fact 검증을 위한 데이터 셋 구축

### ■ TopicD: 토픽 추출

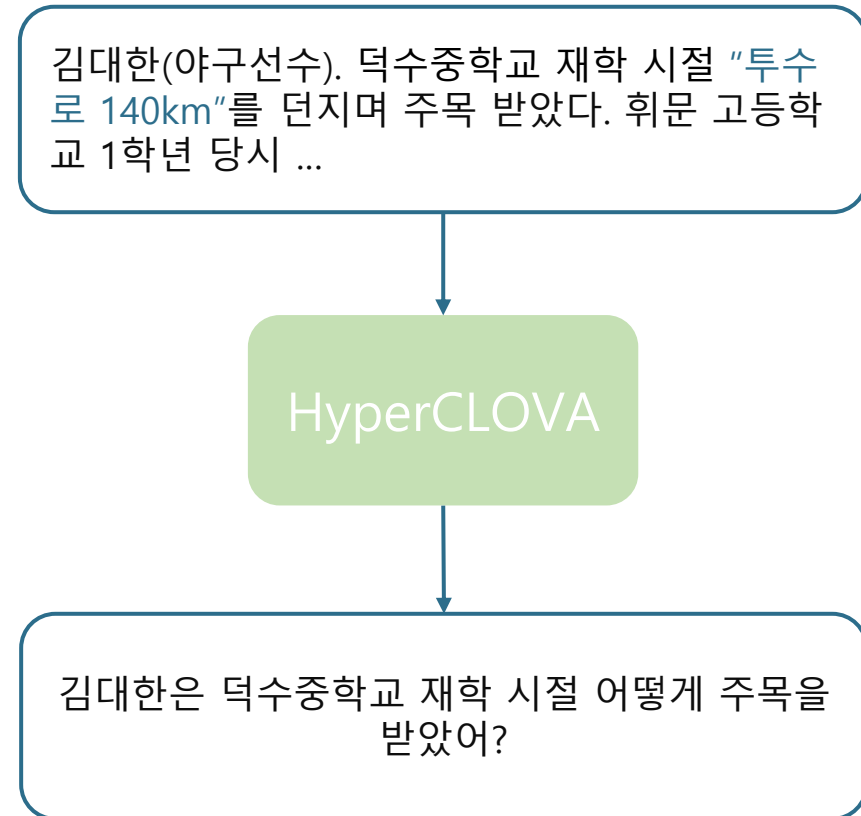
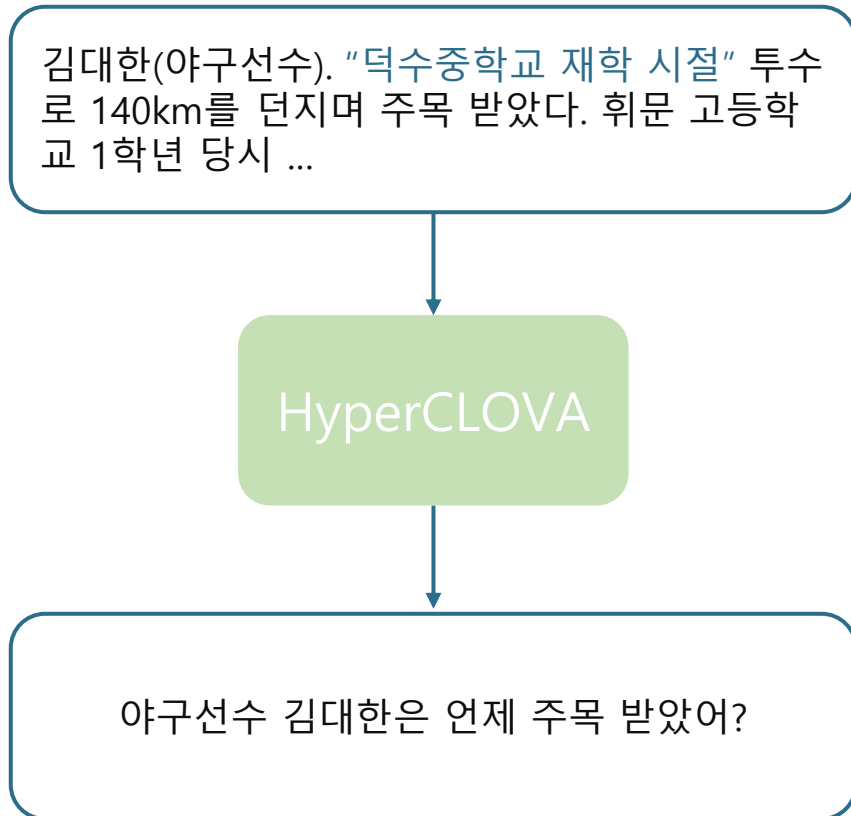
- 주어진 증거 문장 내에서 질문의 대상이 될 만한 스팬을 얻어내는 것이 목적임



## 2. Fact 검증을 위한 데이터 셋 구축

### ■ TopicD: 질문 생성

- 토픽 추출 단계에서 얻어진 토픽을 이용해 질문을 생성





## 2. Fact 검증을 위한 데이터 셋 구축

### ■ TopicD: 질문 생성

- 토픽 추출 단계에서 얻어진 토픽을 이용해 질문을 생성

생성된 질문	Context
미네소타 와일드는 어느 나라 팀이야?	미네소타 와일드(Minnesota Wild)는 "미국 미네소타주 세인트폴에 소재한 프로 아이스하키 팀"이다..
미네소타 와일드는 어느 리그에 속해있어?	이 팀은 1997년에 창단되었으며, 현재 "NHL 서부 콘퍼런스 센트럴 디비전"에 속해 있다.
판의 경계에는 무슨 특징이 있어?	판의 경계에서는 "지진과 화산활동이 잦으며, 대부분의 화산은 판의 경계에 따라 있다"
판의 경계는 무슨 종류가 있어?	판의 경계는 "발산형경계와 수렴형경계, 보존형경계의 3가지"로 구분할 수 있다.

## 2. Fact 검증을 위한 데이터 셋 구축

### ■ TopicD: 답변 생성

- HyperCLOVA 모델을 이용하여 입력 증거 문장과 질문으로부터 증거 문장에 기반한 SUPPORTED 답변을 얻어냄

질문	증거 문장	생성된 답변
미네소타 와일드는 어느 나라 팀이야?	미네소타 와일드(Minnesota Wild)는 미국 미네소타주 세인트폴에 소재한 프로 아이스하키 팀이다..	미네소타 와일드(Minnesota Wild)는 미국 미네소타주 세인트폴에 소재한 프로 아이스하키 팀입니다.
근은 몇 g이야?	근(斤)은 질량과 무게의 단위이며, 1근은 16냥 또는 160돈이며, 미터법으로는 약 600g이다.	근은 160돈 또는 600g 입니다.
드라마 흑기사는 언제 방송했어?	《흑기사》는 2017년 12월 6일부터 2018년 2월 8일까지 KBS 2TV에서 방송된 수목드라마이다.	《흑기사》는 2017년 12월 6일부터 2018년 2월 8일까지 KBS 2TV에서 방송되었습니다.
사포딜라의 원산지는 어디야?	원산지는 과테말라, 니카라과, 멕시코, 벨리즈 등 북·중미 지역이다.	사포딜라의 원산지는 과테말라, 니카라과, 멕시코, 벨리즈 등 북·중미 지역입니다.

## 2. Fact 검증을 위한 데이터 셋 구축

### ■ TopicD: 답변 변형

- 답변 생성 단계에서 얻어진 지문에 근거한 SUPPORTED 답변으로부터 REFUTED 답변을 얻어냄

기존 답변	변경된 답변
카를로스 사인스는 스페인의 자동차 경주 선수입니다.	카를로스 사인스는 <b>미국</b> 의 자동차 경주 선수입니다.
Black Eyes는 대한민국의 걸 그룹 티아라의 세 번째 미니 앨범입니다.	Black Eyes는 대한민국의 걸 그룹 티아라의 세 번째 미니 앨범입니다.
오후의 재발견은 2006년 4월 24일부터 2009년 4월 19일까지 춘천MBC FM4U(FM 94.5MHz)에서 매일 오후 4시부터 저녁 6시까지 방송했던 대중음악 전문 라디오 프로그램입니다.	오후의 재발견은 <b>1999년 8월 12일부터 2010년 6월 15일까지</b> 춘천MBC FM4U(FM 94.5MHz)에서 매일 오후 4시부터 저녁 6시까지 방송했던 대중음악 전문 라디오 프로그램입니다.
흥부전은 조선 시대의 작가 미상의 고전 소설로 빈부 격차에 대한 비판 내용을 담고 있으며, 한국에서 널리 알려진 이야기입니다.	흥부전은 <b>삼국</b> 시대의 작가 미상의 고전 소설로 빈부 격차에 대한 비판 내용을 담고 있으며, 한국에서 널리 알려진 이야기입니다.

## 2. Fact 검증을 위한 데이터 셋 구축

### ■ TopicD: 답변 변형

- 답변 생성 단계에서 얻어진 지문에 근거한 SUPPORTED 답변으로부터 REFUTED 답변을 얻어냄
- 주어를 제거한 후 얻어낸 REFUTED 답변

기존 답변	변경된 답변
카를로스 사인스는 스페인의 자동차 경주 선수입니다.	브라질 축구 선수입니다.
Black Eyes는 대한민국의 걸 그룹 티아라의 세 번째 미니 앨범입니다.	일본 걸 그룹 베리즈코보의 네 번째 싱글 앨범입니다.
오후의 재발견은 2006년 4월 24일부터 2009년 4월 19일까지 춘천MBC FM4U(FM 94.5MHz)에서 매일 오후 4시부터 저녁 6시까지 방송했던 대중음악 전문 라디오 프로그램입니다.	아침 10시에 방송되는 교양 프로그램입니다.
흥부전은 조선 시대의 작가 미상의 고전 소설로 빈부 격차에 대한 비판 내용을 담고 있으며, 한국에서 널리 알려진 이야기입니다.	미국에서 만들어진 뮤지컬이며, 국내에는 없는 것으로 보입니다.

## 2. Fact 검증을 위한 데이터 셋 구축

### ■ SumD

- 모델의 multi-hop 추론 능력 향상을 위해 생성한 데이터 셋
- TopicD 에 비해 절차가 간단하며 [2]의 BART 모델과 HyperCLOVA 모델을 이용하여 자동 구축함

증거 문서	이이익: 이이익 1868년 4월 26일 ~ 1935년 3월 12일)은 대한제국 관료 출신 독립운동가이다. 호는 산남. 경상북도 칠곡에서 출생하였고 지난날 한때 경상북도 달성에서 잠시 유아기를 보낸 적이 있으며 ...
생성된 질문	이이익에 대해 알려줘
SUPPRTED 답변	대한제국 관료 출신 독립운동가 이이익은 1881년 인동 장씨 문중의 6년 연상녀 장진평과 결혼 후 1892년 음서로써 관직에 천거되어 1903년에서 1909년까지 경상북도 칠곡군 인동 태수 직을 지냈습니다.
REFUTED 답변	대한제국 관료 출신 독립운동가 이이익은 김응하의 아들이자 대한제국의 무관 이자 군인 이었습니다.

### 3. 실험

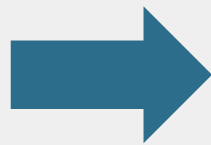
#### ■ 증거 문서 검색 및 증거 문장 선택

- 주어진 주장이 사실인지 아닌지를 판단하기 위해서는 주장에 대한 증거(evidence)가 뒷받침 되어야 함
- 주장에 대한 증거 문서를 검색하기 위해 검색 모델인 REALM[3] 을 이용함
- 
- 검색된 증거 문서 중 적절한 증거 문장 선택 과정에는 TF-IDF 를 이용함



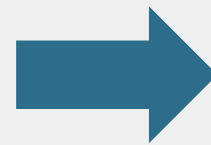
전체 문서 셋

REALM



검색된 증거 문서

TF-IDF



선택된 후보 증거 문장들

# 3. 실험

## ■ 실험 세팅

- 세 가지 방법을 이용하여 자동 생성된 데이터 셋을 이용하여 모델 학습을 진행 (TopicE, TopicD, SumD)

	<b>TopicE</b>	<b>TopicD</b>	<b>SumD</b>
학습 셋 (Train set)	4,689	7,190	6,100
개발 셋 (Dev set)	400	500	500

- 데이터 셋 제작에는 39B 크기의 HyperCLOVA 모델을 사용함
- 최근 Fact 검증 태스크에서 좋은 성능을 거둔 [4]의 KGAT 모델, [5]의 GERT 모델과 [6]의 MLA 모델을 사용하여 학습 및 평가를 진행함

[4] Fine-grained fact verification with kernel graph attention network, [Liu et al., 19]

[5] Gear: Graph-based evidence aggregation and reasoning for fact verification, [Zhou et al., 19]

[6] A multi-level attention model for evidence-based fact checking, [Kruengkrai et al., 21]

# 3. 실험

## ■ 실험 세팅

- 평가 셋은 네이버로부터 제공받은 평가 셋을 이용함 (네이버 지식백과 평가 셋)

사용자 원질의	질의 정규화	지식백과 내 근거 문장	지식백과 링크	39B_기획셋학습	신뢰도	39B기획 평가 (1/0/-1)
머리카락은 잘라도 왜 계속 자라	-	머리카락을 뽑으면 모근만 뾰허	<a href="https://terms.n">https://terms.n</a>	머리카락은 모낭이라는 곳에서 만들어지는데 머리카락이 계속 자라는 이유는 모낭 안에 새로운	(신뢰도 높음)	1
모낭에 대해서 알려줘	-	그건 바로 작은 주머니처럼 생긴	<a href="https://terms.n">https://terms.n</a>	모낭은 모근을 감싸고 털에 영양을 공급하는 주머니입니다.	(신뢰도 중간)	1
모근이 뭐야	-	모근 포유류의 피부에 나는 털의	<a href="https://terms.n">https://terms.n</a>	모근은 머리카락의 뿌리를 말해요.	(신뢰도 높음)	1
머리카락이 자라는 속도 알려줘	-	하루에 0.2~0.4mm, 한 달에 약	<a href="https://terms.n">https://terms.n</a>	하루에 0.2~0.4mm, 한 달에 약 1cm정도 자랍니다.	(신뢰도 중간)	1
진달래 먹을수 있어	-	바로 먹을 수 있는 꽃. 식품으로 적	<a href="https://terms.n">https://terms.n</a>	네 화전이나 떡에 넣어 만들어 먹을 수 있습니다.	(신뢰도 낮음)	1
진달래로 뭐 만들어먹어	-	관상용으로 심으며 꽃을 화전이나	<a href="https://terms.n">https://terms.n</a>	조상들은 진달래로 술, 떡, 화전 등 다양한 음식을 만들었습니다.	(신뢰도 낮음)	1
화전이 뭐야	-	참쌀가루를 반죽하여 기름에 지진	<a href="https://terms.n">https://terms.n</a>	진달래꽃을 녹두 가루와 반죽하여 팬에 지진 떡입니다.	(신뢰도 중간)	0
철쭉도 먹을수 있어	-	꽃은 독성이 있어 진달래와 달리	<a href="https://terms.n">https://terms.n</a>	아니오 철쭉은 독이 있어 먹을 수 없습니다.	(신뢰도 낮음)	1
독성이라는게 뭐야	-	독소 또는 화합물이 생물에게 피	<a href="https://terms.n">https://terms.n</a>	독성은 어떤 물질이 생체에 해를 끼치며, 그 물질을 섭취했을 때 건강에 해를 끼칠 수 있는 성질	(신뢰도 중간)	1
임진왜란이 뭐야	-	조선 선조 25년(1592년)부터 3	<a href="https://terms.n">https://terms.n</a>	1592년, 일본의 도요토미 히데요시가 15만 왜군을 보내 조선을 침략한 걸 말해요	(신뢰도 중간)	1
어느 왕때 일어난 전쟁이야	임진왜란은 어느 왕때 일어난 전쟁이야	조선 선조 25년(1592년)부터 3	<a href="https://terms.n">https://terms.n</a>	임진왜란은 조선의 선조 때 일어난 전쟁입니다.	(신뢰도 높음)	1
선조는 조선 몇대 왕이었어	-	조선의 제14대 임금이다. 처음에	<a href="https://terms.n">https://terms.n</a>	선조는 제14대 왕이었습니다.	(신뢰도 낮음)	1
우리가 이겼어	임진왜란은 우리가 이겼어	일본은 명나라를 정복하려 가는	<a href="https://terms.n">https://terms.n</a>	네 이순신 장군의 활약으로 우리가 승리하였습니다.	(신뢰도 낮음)	1
이순신 장군은 누구야	-	임진왜란이 일어나기 바로 전 해	<a href="https://terms.n">https://terms.n</a>	이순신 장군은 조선 중기의 무신입니다	(신뢰도 높음)	1
어떻게 이긴거야	이순신 장군은 어떻게 이긴거야	거북선은 임진왜란 때 이순신 장	<a href="https://terms.n">https://terms.n</a>	뛰어난 전략가이자 전황을 정확히 분석하여 적절한 전술을 구사하였으며, 거북선과 같은 뛰어난	(신뢰도 중간)	1
2개월된 강아지는 잠을 얼마나 자	-	강아지는 약 17~18시간 정도 잠	<a href="https://terms.n">https://terms.n</a>	강아지들은 생후 6~8개월 전까지 하루의 대부분을 잠으로 보냅니다	(신뢰도 높음)	1
잠을 많이 못자면 어떻게 될까	강아지들이 잠을 많이 못자면 어떻게 될까	사람과 마찬가지로 반려견도 필요	<a href="https://terms.n">https://terms.n</a>	하루에 10시간 이상 잠을 자지 못하면 면역력이 떨어지고 질병에 걸릴 확률이 높아집니다.	(신뢰도 높음)	1
강아지는 어디에서 자는게 좋아	-	반려견은 외부 자극에 쉽게 잠을	<a href="https://terms.n">https://terms.n</a>	강아지들은 보통 지붕이 있는 집에서 잠을 잡니다	(신뢰도 높음)	1



### 3. 실험

#### ■ 실험 결과

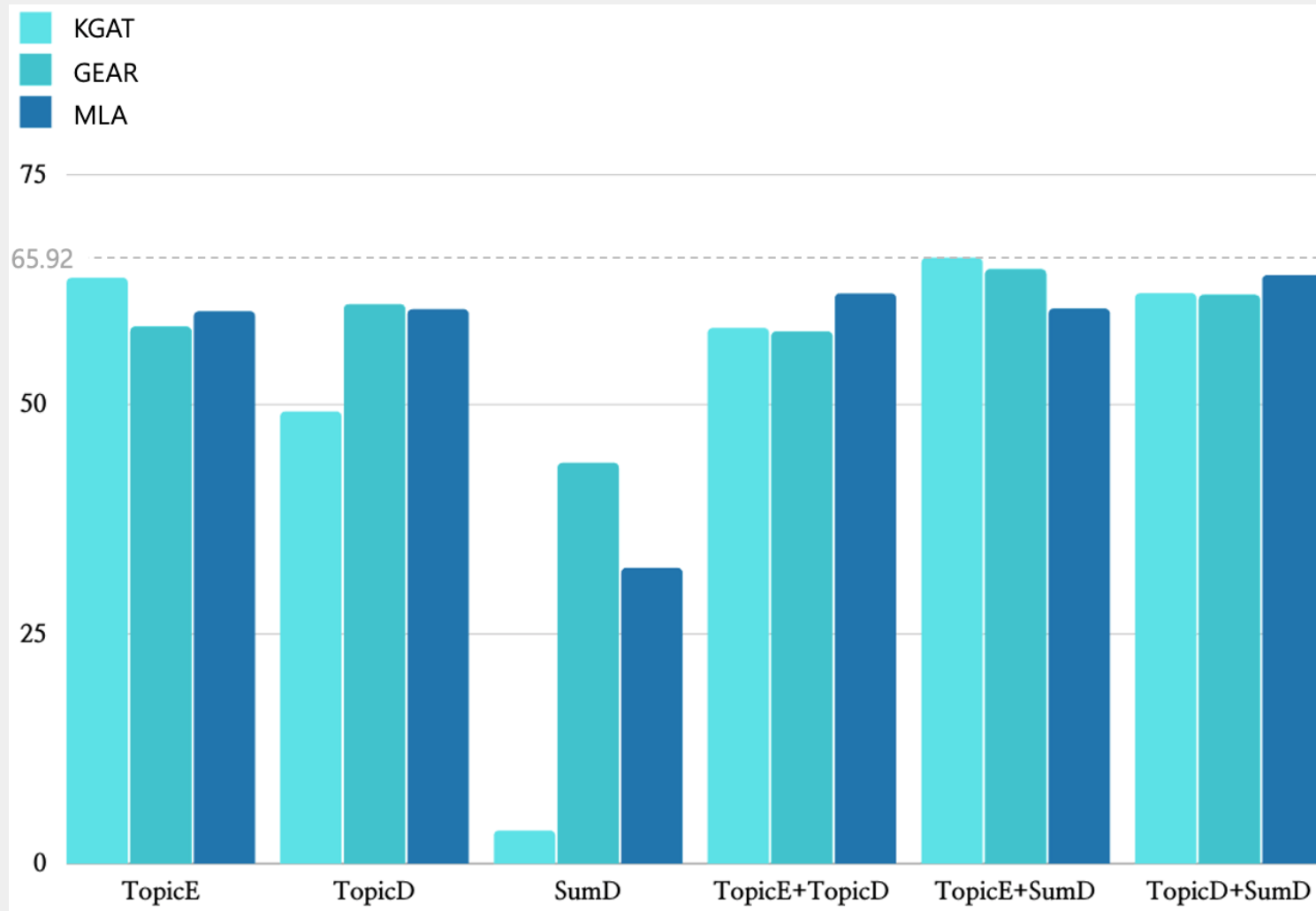
- 실험 결과, KGAT 모델이 F1 score 65.93% 로 가장 높은 성능을 달성함

모델	Precision	Recall	F1 score	Accuracy
KGAT	82.19%	55.04%	<b>65.93%</b>	<b>70.89%</b>
GEAR	71.26%	56.88%	63.26%	66.19%
MLA	70.00%	57.79%	63.31%	65.72%

# 3. 실험

## ■ 실험 결과

- 각 데이터 셋이 모델의 성능에 미치는 영향



## 4. 결론

- 최근 활발히 연구되고 있는 Fact 검증 작업을 수행하기 위해 HyperCLOVA 모델을 이용하여 사실 확인 데이터 셋을 자동으로 구축하여 최근 좋은 성능을 보이고 있는 Fact 검증 모델들을 학습 및 그 성능을 측정함
- 최종적으로 KGAT 모델의 성능이 F1 score 65.93%로 가장 높은 성능을 달성함
- 기존 Fact 검증 데이터와 달리 주장(Claim)을 질문과 답변의 형태로 나누었기 때문에 드러난 한계점은 질문과 답변의 연관성을 고려한 모델의 추론 능력을 동시에 요구한다는 것 이었음
- 즉, 답변이 질문에서 요구하는 명확한 답변은 아니지만 증거 문서에 기반한 사실인 경우 모델은 이를 SUPPORTED 라고 예측하는 경우가 다수 존재함
- 결과적으로 이러한 답변과 질문 사이의 연관성을 고려한 Fact 검증 모델의 추론은 앞으로 더 연구가 필요한 주제라고 생각됨

감사합니다