(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0131251
(43) 공개일자 2024년08월30일

(51) 국제특허분류(Int. Cl.)
*G06F 16/332* (2019.01) *G06F 16/33* (2019.01)
*G06F 16/34* (2019.01) *G06N 3/0475* (2023.01)
*G06N 3/096* (2023.01) *G06Q 50/20* (2012.01)
(52) CPC특허분류
*G06F 16/3329* (2019.01)
*G06F 16/3344* (2019.01)
(21) 출원번호 10-2024-0009742
(22) 출원일자 2024년01월22일
심사청구일자 2024년01월22일
(30) 우선권주장
1020230024355 2023년02월23일 대한민국(KR)

(71) 출원인
**고려대학교 산학협력단**
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
**임희석**
경기도 수원시 영통구 영통로 460, 304동 1704호
**어수경**
경상남도 합천군 삼가면 금리4길 17
*(뒷면에 계속)*
(74) 대리인
**김등용**

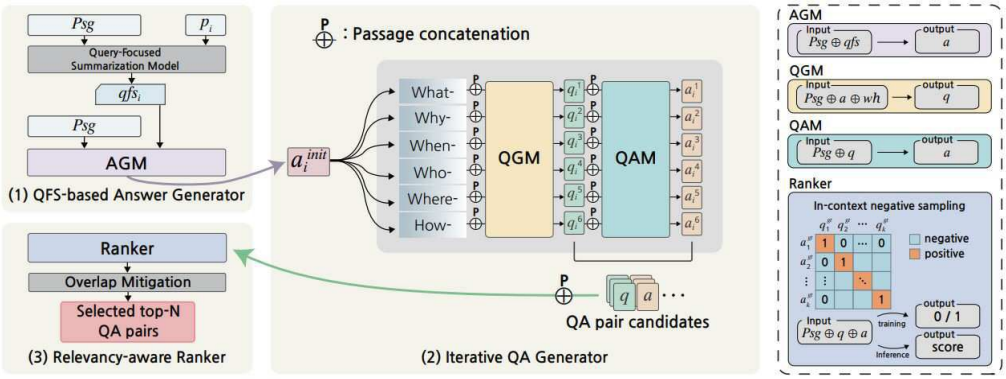
전체 청구항 수 : 총 9 항

(54) 발명의 명칭 **유형 다양성을 고려한 질의응답쌍 생성 장치 및 방법**

## (57) 요 약

유형 다양성을 고려한 교육용 질의응답쌍 생성 장치 및 방법이 개시된다. 상기 질의응답쌍 생성 방법은 컴퓨팅 장치에 의해 수행되고, 단락(passage)에 대한 질의집중요약(Query-Focused Summarization, QFS)을 생성하는 단계, 상기 단락 및 상기 질의집중요약을 기초로, 초기 정답을 생성하는 단계, 상기 초기 정답, 상기 단락, 및 의문사를 기초로, 상기 초기 정답에 대응하는 질의를 생성하는 단계, 상기 질의 및 상기 단락을 기초로, 상기 질의에 대응하는 응답을 생성하여, 질의응답쌍을 생성하는 단계, 및 상기 질의응답쌍 중에서, 적어도 하나의 질의응답쌍을 선택하여 최종 질의응답쌍을 도출하는 단계를 포함한다.

*대 표 도* - 도1



(52) CPC특허분류
*G06F 16/345* (2019.01)
*G06N 3/0475* (2023.01)
*G06N 3/096* (2023.01)
*G06Q 50/20* (2013.01)
(72) 발명자
**문현석**
서울특별시 강남구 자곡로 260, 422동 903호
**김진성**
서울특별시 성북구  안암로9가길 85, 204호
**허윤아**
경기도 수원시 영통구 영통로 111, 306동 1106호
**김정욱**
강원도 강릉시 선수촌로 113, 805동 904호

이 발명을 지원한 국가연구개발사업
과제고유번호 1711159969
과제번호 2018-0-01405-005
부처명 과학기술정보통신부
과제관리(전문)기관명 정보통신기획평가원
연구사업명 정보통신방송혁신인재양성
연구과제명 Human-inspired 복합지능 원천기술 연구 및 인력양성
과제수행기관명 고려대학교산학협력단
연구기간 2023.01.01 ~ 2023.12.31
이 발명을 지원한 국가연구개발사업
과제고유번호 1345349433
과제번호 2021R1A6A1A03045425
부처명 교육부
과제관리(전문)기관명 한국연구재단
연구사업명 이공학학술연구기반구축
연구과제명 Human-inspired AI 연구소
과제수행기관명 고려대학교
연구기간 2022.03.01 ~ 2023.02.28

# 명 세 서

## 청구범위

### 청구항 1

컴퓨팅 장치에 의해 수행되는 질의응답쌍(Question-Answer Pair, QA쌍) 생성 방법에 있어서,

단락(passage)에 대한 질의집중요약(Query-Focused Summarization, QFS)을 생성하는 단계;

상기 단락 및 상기 질의집중요약을 기초로, 초기 정답을 생성하는 단계;

상기 초기 정답, 상기 단락, 및 의문사를 기초로, 상기 초기 정답에 대응하는 질의를 생성하는 단계;

상기 질의 및 상기 단락을 기초로, 상기 질의에 대응하는 응답을 생성하여, 질의응답쌍을 생성하는 단계; 및

상기 질의응답쌍 중에서, 적어도 하나의 질의응답쌍을 선택하여 최종 질의응답쌍을 도출하는 단계를 포함하는 질의응답쌍 생성 방법.

### 청구항 2

제1항에 있어서,

상기 QFS를 생성하는 단계는, 상기 단락에 포함되는 문장의 개수만큼의 질의집중요약을 생성하는,

질의응답쌍 생성 방법.

### 청구항 3

제1항에 있어서,

상기 초기 정답을 생성하는 단계는, 단락 및 질의집중요약을 입력받아 초기 정답을 생성하도록 사전학습된 정답 생성 모델에 상기 단락 및 상기 질의집중요약을 입력하여 상기 초기 정답을 생성하는,

질의응답쌍 생성 방법.

### 청구항 4

제1항에 있어서, 상기 질의를 생성하는 단계는, 초기 정답, 단락, 및 의문사를 입력받아 질의를 생성하도록 사전학습된 질의 생성 모델에 상기 초기 정답, 상기 단락, 및 상기 의문사를 입력하여 상기 질의를 생성하고,

상기 의문사는 What, Why, When, Who, Where, 및 How인,

질의응답쌍 생성 방법.

### 청구항 5

제1항에 있어서,

상기 질의응답쌍을 생성하는 단계는, 질의 및 단락을 입력받아 응답을 생성하도록 사전학습된 질의응답 모델에 상기 질의 및 상기 단락을 입력하여 상기 응답을 생성하는,

질의응답쌍 생성 방법.

### 청구항 6

제1항에 있어서,

상기 최종 질의응답쌍을 도출하는 단계는,

상기 질의응답쌍의 순위 점수를 도출하는 단계; 및

상기 순위 점수가 가장 높은 질의응답쌍을 선택하는 단계를 포함하는,

질의응답쌍 생성 방법.

**청구항 7**

제6항에 있어서,

상기 순위 점수를 도출하는 단계는,

입력되는 질의응답쌍이 옳은 예시인지 틀린 예시인지에 대해 이진 분류하도록 사전학습된 순위 생성 모델에 상기 질의응답쌍을 입력하여 상기 순위 점수를 도출하는,

질의응답쌍 생성 방법.

**청구항 8**

제7항에 있어서,

상기 순위 점수를 도출하는 단계는, 상기 질의응답쌍이 옳은 예시로 분류될 확률을 나타내는,

질의응답쌍 생성 방법.

**청구항 9**

제6항에 있어서,

상기 순위 점수를 도출하는 단계는,

상기 순위 점수가 제일 높은 질의응답쌍과 나머지 질의응답쌍들 간의 Rouge-L 스코어를 측정하는 단계;

상기 순위 점수에서, 상기 Rouge-L 스코어와 상기 순위 점수의 절대값의 곱을 차감하여, 각 질의응답쌍의 조정된 순위 점수를 도출하는 단계; 및

상기 조정된 순위 점수가 가장 높은 질의응답쌍을 선택하는 단계를 포함하는,

질의응답쌍 생성 방법.

## *발명의 설명*

### *기 술 분 야*

[0001] 본 발명은 인공지능 기술 중 자연어 처리(Natural Language Processing, NLP)에 관한 것으로, 보다 상세하게는 주어진 단락(passage)에 대한 질의응답쌍을 생성하는 장치 및 방법에 관한 것이다.

### *배 경 기 술*

[0002] QA쌍 생성(Question-Answer Pair Generation, QAG) 분야에서의 최근의 발전으로, 이 기술을 교육 분야에 적용하기에 이르렀다. 그러나, QA 유형(QA types)의 다양성은, 그 기여에도 불구하고, 어린이들의 포괄적인 학습과 평가에 여전히 도전적인 과제이다. 본 발명에서는, 다양한 의문문(interrogative sentences)과 암시적 및/또는 명시적 응답을 생성함으로써, QA 유형 다양성을 향상시키는 QAG 프레임워크를 제안한다. 제안하는 프레임워크는, 질의집중요약 기반 정답 생성기(Query-Focused Summarization-based answer generator), 순환적인 QA 생성기(inerative QA generator), 및 관련성 중심의 순위 생성기(relevancy-aware ranker)를 포함한다.

[0003] 두 생성기(정답 생성기와 QA 생성기)는, 다양한 유형을 포괄하면서, 후보의 수를 확장하는 것을 목표로 한다. 맥락내 부정 샘플(in-context negative samples)을 기반으로 학습된 순위 생성기는, 순위 점수(ranking score)에 기초하여, 상위 N 출력(top-N outputs)을 명확하게 한다. 평가와 상세한 분석은, 제안하는 접근 방식이 이전 최고 성능 결과를 큰 폭으로 능가하며, 다양성과 품질을 향상시키는 것을 입증한다. 제안하는 태스크-중심 프로세스(task-oriented processes)는 실제 필요(real-world demand)에 부합하며, 이는 제안하는 시스템의 높은 적용 가능성을 강조한다.

### *선행기술문헌*

### *비특허문헌*

[0004] (비특허문헌 0001) Jesse Vig, Alexander Fabbri, Wojciech Kryscinski, Chien-Sheng Wu, and Wenhao Liu. 2022. Exploring neural models for query-focused summarization. In Findings of the Association for Computational Linguistics: NAACL 2022, pages 1455-1468, Seattle, United States. Association for Computational Linguistics.

## 발명의 내용

### 해결하려는 과제

[0005] 본 발명이 이루고자 하는 기술적인 과제는 유형 다양성을 고려한 질의응답쌍 생성 장치 및 방법을 제공하는 것이다.

### 과제의 해결 수단

[0006] 본 발명의 일 실시예에 따른, 질의응답쌍 생성 방법은 컴퓨팅 장치에 의해 수행되는 질의응답쌍(Question-Answer Pair, QA쌍) 생성 방법으로, 단락(passage)에 대한 질의집중요약(Query-Focused Summarization, QFS)을 생성하는 단계, 상기 단락 및 상기 질의집중요약을 기초로, 초기 정답을 생성하는 단계, 상기 초기 정답, 상기 단락, 및 의문사를 기초로, 상기 초기 정답에 대응하는 질의를 생성하는 단계, 상기 질의 및 상기 단락을 기초로, 상기 질의에 대응하는 응답을 생성하여, 질의응답쌍을 생성하는 단계, 및 상기 질의응답쌍 중에서, 적어도 하나의 질의응답쌍을 선택하여 최종 질의응답쌍을 도출하는 단계를 포함한다.

### 발명의 효과

[0007] 본 발명의 실시예들에 의할 경우, 가능한 다양한 유형의 질의응답쌍을 생성하며 이중 우수한 품질이 예측되는 질의응답쌍만을 최종 결과로 채택하는 방식을 활용한다. 또한, 이를 통해 기존 QAG(QA pair generation) 연구에서의 질의응답쌍 생성 퀄리티를 높이면서 동시에 질의 및 응답 유형의 다양성을 크게 증가시킬 수 있다.

### 도면의 간단한 설명

[0008] 본 발명의 상세한 설명에서 인용되는 도면을 보다 충분히 이해하기 위하여 각 도면의 상세한 설명이 제공된다.

도 1은 본 발명에서 제안하는 QAG(질의응답쌍 생성) 프레임워크의 전체 아키텍처를 도시한다.

도 2는 본 발명의 일 실시예에 따른, 질의응답쌍 생성 방법을 설명하기 위한 흐름도이다.

### 발명을 실시하기 위한 구체적인 내용

[0009] 본 명세서에 개시되어 있는 본 발명의 개념에 따른 실시예들에 대해서 특정한 구조적 또는 기능적 설명들은 단지 본 발명의 개념에 따른 실시예들을 설명하기 위한 목적으로 예시된 것으로서, 본 발명의 개념에 따른 실시예들은 다양한 형태들로 실시될 수 있으며 본 명세서에 설명된 실시예들에 한정되지 않는다.

[0010] 본 발명의 개념에 따른 실시예들은 다양한 변경들을 가할 수 있고 여러 가지 형태들을 가질 수 있으므로 실시예들을 도면에 예시하고 본 명세서에서 상세하게 설명하고자 한다. 그러나, 이는 본 발명의 개념에 따른 실시예들을 특정한 개시 형태들에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물, 또는 대체물을 포함한다.

[0011] 제1 또는 제2 등의 용어는 다양한 구성 요소들을 설명하는데 사용될 수 있지만, 상기 구성 요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성 요소를 다른 구성 요소로부터 구별하는 목적으로만, 예컨대 본 발명의 개념에 따른 권리 범위로부터 벗어나지 않은 채, 제1 구성 요소는 제2 구성 요소로 명명될 수 있고 유사하게 제2 구성 요소는 제1 구성 요소로도 명명될 수 있다.

[0012] 어떤 구성 요소가 다른 구성 요소에 "연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 그 다른 구성 요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성 요소가 존재할 수도 있다고 이해되어야 할 것이다. 반면에, 어떤 구성 요소가 다른 구성 요소에 "직접 연결되어" 있다거나 "직접 접속되어" 있다고 언급된 때에는 중간에 다른 구성 요소가 존재하지 않는 것으로 이해되어야 할 것이다. 구성 요소들 간의 관계를 설명하는 다른 표현들, 즉 "~사이에"와 "바로 ~사이에" 또는 "~에 이웃하는"과 "~에 직접 이웃하는" 등

도 마찬가지로 해석되어야 한다.

[0013] 본 명세서에서 사용한 용어는 단지 특정한 실시예를 설명하기 위해 사용된 것으로서, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "포함하다" 또는 "가지다" 등의 용어는 본 명세서에 기재된 특징, 숫자, 단계, 동작, 구성 요소, 부분품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성 요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[0014] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가진다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥상 가지는 의미와 일치하는 의미를 갖는 것으로 해석되어야 하며, 본 명세서에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.

[0015] 이하, 본 명세서에 첨부된 도면들을 참조하여 본 발명의 실시예들을 상세히 설명한다. 그러나, 특허출원의 범위가 이러한 실시예들에 의해 제한되거나 한정되는 것은 아니다. 각 도면에 제시된 동일한 참조 부호는 동일한 부재를 나타낸다.

[0016] 본 발명에서 제안하는 QAG(질의응답쌍 생성) 프레임워크는 세 가지의 태스크 중심 프로세스(task-oriented processes), 즉 질의집중요약 기반 정답 생성기(Query-Focused Summarization-based answer generator), 순환적인 질의응답 생성기(iterative QA generator), 및 관련성 중심의 순위 생성기(relevancy-aware ranker)를 포함한다. 두 생성기(정답 생성기와 질의응답 생성기)의 주요 목표는 다양한 질의 및 응답 유형을 포함하는 QA쌍 후보(QA pair candidates)를 확장하는 것이다. 순위 생성기는 QA쌍 후보를 스코어링함으로써 최종 출력을 결정하는 것을 목표로 한다. 제안하는 프레임워크의 전체 QAG 아키텍처는 도 1에 도시되어 있다.

[0017] **질의집중요약 기반 정답 생성기(QFS-based Answer Generator)**

[0018] 초기 정답 생성 과정에서, 주어진 문장과 관련된 중요한 정보(salient information)를 포착하기 위해 질의집중요약(query-focused summarization, QFS)을 활용한다. QFS 모델이, 관련 키(relevant key)를 참조함으로써, 주어진 단락(passage)의 질의집중요약(QFS)을 생성한다. 본 발명에서 제안하는 질의집중요약 기반 정답 생성기는 Vig 등의 질의집중요약 생성 모델(또는 질의집중요약 생성 기법)을 활용할 수 있다(Jesse Vig, Alexander Fabbri, Wojciech Kryscinski, Chien-Sheng Wu, and Wenhao Liu. 2022. Exploring neural models for query-focused summarization. In Findings of the Association for Computational Linguistics: NAACL 2022, pages 1455-1468, Seattle, United States. Association for Computational Linguistics.). 또한, 이에 대한 상세한 설명은 상기 논문이 참조될 수 있다. 이후, 생성된 요약은, 명시적 및/또는 암시적 정답을 출력하기 위해, 생성적(generative) 정답 생성 모델(Answer Generation Model, AGM)로 입력된다.

[0019] *Psg*를 n 문장들($p_1$, …, $p_n$)로 구성된 단락(passage)이라 하고, 대응하는 GT(ground-truth) QA쌍을 $(Q^{\text{gt}}, A^{\text{gt}}) = \{(q_j^{\text{gt}}, a_j^{\text{gt}})\}_{j=1}^{m}$라 하자. 먼저, 사전학습된(pre-trained) QFS 모델(*QFS*)을 이용하여, *Psg*의 질의집중요약 $qfs_j^{\text{gt}} = QFS(Psg, q_j^{\text{gt}})$을 생성한다. 그런 다음, *Psg*와 $qfs_j^{\text{gt}}$의 연결된 입력(concatenated input)을 시퀀스 투 시퀀스 방식(sequence-to-sequence manner)으로 $\theta_{\text{AGM}}$으로 명명된 AGM을 학습시킨다. 각 *Psg*에 대한 손실함수(loss function)는 수학식 1과 같이 추정된다.

[0020] [수학식 1]

[0021] $$L_{\text{AGM}} = -\sum_{(q_j^{\text{gt}}, a_j^{\text{gt}}) \in (Q^{\text{gt}}, A^{\text{gt}})} \mathbf{E}_{\theta_{\text{AGM}}}(a_j^{\text{gt}} \mid Psg, qfs_j^{\text{gt}})$$

[0022] 추론 단계(inference phase)에서는, *Psg* 내의 각 문장 $p_i$에 대해, $qfs_i = QFS(Psg, p_i)$를 생성한다. 그런 다음, AGM은 해당 $qfs_i$에 대한 단일의 초기 정답(single initial answer) $a_i^{init}$를 생성한다. 결과적인 정답 집합 $A^{init}$은, 정답은 단락 내의 모든 문장에 대해 생성되기 때문에, n개의 정답을 포함한다. $A^{init}$는 수학식 2와 같이 표현된다.

[0023] [수학식 2]

[0024] $$A^{init} = \{\theta_{\text{AGM}}(Psg, qfs_i) \mid p_i \in Psg\}$$

[0025] **순환적인 질의응답 생성기(Iterative QA Generator)**

[0026] 초기 정답 집합 $A^{init}$이 생성된 후, 다음 스텝은, 질의 유형 다양성(question type diversity)을 반영하기 위해, QA쌍 후보(QA pair candidates)를 확장하는 것이다. 이를 달성하기 위해, $\theta_{\text{QGM}}$으로 표기되는 의문사 지시 질의 생성 모델(interrogative word-indicated question generation model, QGM)과 $\theta_{\text{QAM}}$으로 표기되는 생성적 질의 응답 모델(generative question-answering model, QAM)을 제안한다. QGM과 QAM은, QA쌍 후보의 집합을 생성하기 위해, 초기 정답을 기반으로 순차적으로 실행된다. 다음 단락에서는 각 모델의 교육 및 추론 과정을 설명한다.

[0027] **의문사 지시 질의 생성 모델(Interrogative word-indicated QGM)** 정답 및 단락을 참조함으로써, 질의를 생성하기 위해, GT QA쌍 세트로 QGM을 학습한다. 학습 단계(training phase)에서 의문사(interrogative words)를 포함시키면, 추론 중에 원하는 의문 유형(interrogative type)을 따르는 제어 가능한 질의 생성이 가능해 진다.

[0028] GT QA쌍 집합 내의 각 $q_j^{\text{gt}}$의 의문사(interrogative word)를 $wh_j^{\text{gt}}$로 표기한다. 설정(setting)에서, $wh$는 의문사 집합(interrogative word set) $WH=\{\text{Who, When, What, Where, Why, How}\}$의 요소(element)이다. $\theta_{\text{QGM}}$은, $Psg$, $a_j^{\text{gt}}$, 및 $wh_j^{\text{gt}}$의 결합된 입력(concatenated input)을 입력함으로써, 질의 $q_j^{\text{gt}}$를 생성하도록 학습된다. 학습은 시퀀스 투 시퀀스 방식으로 수행되고 수학식 3의 손실 함수를 이용하여 최적화된다.

[0029] [수학식 3]

[0030] $$L_{\text{QGM}} = -\sum_{(q_j^{\text{gt}}, a_j^{\text{gt}}) \in (Q^{\text{gt}}, A^{\text{gt}})} \mathbf{E}_{\theta_{\text{QGM}}}(q_j^{\text{gt}} \mid Psg, a_j^{\text{gt}}, wh_j^{\text{gt}})$$

[0031] 추론 단계(inference phase)에서는, 다양성을 우선시하며, $WH$ 내의 각 의문사를 표시자(indicator)로 고려함으로써 질의를 생성한다. 첫번째 스텝에서 생성된 각 $a_i^{init} \in A^{init}$과 이의 대응되는 단락 $Psg$에 대하여, $\theta_{\text{QGM}}$은 수학식 4와 같이 표현될 수 있는 QA쌍 집합 $QA^1$를 구성한다.

[0032] [수학식 4]

[0033] $$QA^1 = \{(\theta_{\text{QGM}}(Psg, a_i^{init}, wh), a_i^{init}) \mid wh \in WH,\ a_i^{init} \in A^{init}\}$$

[0034] 이러한 방식으로, 단락과 높은 연관성을 갖는 QA쌍 후보가 생성될 수 있다. 이런 프로세스는 질의 유형의 확장을 촉진하지만, 생성된 모든 질의가 초기 정답과 관련이 있는 것은 아니다.

[0035] **응답 조정(Answer Adjustment)** QA쌍들 사이의 관련성을 고려하기 위해, GT QA쌍 집합으로 학습된 $\theta_{\text{QAM}}$을 통해 응답을 재구성한다. 이 과정은, 특정한 사람을 목적으로 하는 응답에 대해 "How(어떻게)"라는 질문을 하는 것과 같이, 주어진 초기 정답에 부적절한 질의를 링크(linking)하는 것을 피하도록 돕는다. $\theta_{\text{QAM}}$의 학습은 수학식 5의 손실 함수를 최적화함으로써 진행된다.

[0036] [수학식 5]

[0037] $$L_{\text{QAM}} = -\sum_{(q_j^{\text{gt}}, a_j^{\text{gt}}) \in (Q^{\text{gt}}, A^{\text{gt}})} \mathbf{E}_{\theta_{\text{QAM}}}(a_j^{\text{gt}} \mid Psg, q_j^{\text{gt}})$$

[0038] 이어지는 추론 단계에서, $\theta_{\text{QAM}}$을 통해 $QA^1$ 내의 모든 질의에 대한 응답을 조정한다(adjust). 즉, $QA^1$ 내의 모든 질의 각각에 대한 응답이 새롭게 생성될 수 있다. $QA^2$로 표시되는 재구성된 QA쌍 집합은 수학식 6과 같이 표현된다.

[0039] [수학식 6]

[0040] $$QA^2 = \{(q_i^j, \theta_{\text{QAM}}(Psg, q_i^j)) \mid (q_i^j, a_i^{init}) \in QA^1\}$$

[0041] $QA^2$는, 질의 유형의 다양성을 유지하면서, $\text{QAM}$을 통해 쌍들 사이의 관련성이 지도된(supervised), 최종 QA 쌍 후보 집합을 나타낸다.

[0042] **관련성 중심의 순위 생성기(Relevancy-aware Ranker)**

[0043] 관련성 중심의 순위 생성기(또는 관련성 중심의 순위 생성 모델)로, 단락과 QA쌍 사이의 높은 관련성을 나타내는 상위 N개의 랭크된 QA쌍을 선택할 수 있다.

[0044] $\theta_{Rank}$로 표시되는 순위 생성 모델은 각 QA쌍에 대한 관련성 점수(relevance score)를 생성한다. 순위 생성 모델 $\theta_{Rank}$를 학습하기 위해, GT QA쌍 집합 내에서 맥락내 부정 샘플(in-context negative samples)을 수집함으로써, 대조학습 데이터셋(contrastive training dataset)을 구성할 수 있다. 학습 데이터에서, GT QA쌍은 긍정 샘플(positive samples)로 간주되며, 동일한 단락 내의 다른 QA쌍은 부정 샘플로 간주된다. 보다 구체적으로, 단락-질의-응답쌍 중, 단락과 질의는 유지한체 응답을 단락 내 다른 정답으로 대체하거나, 단락과 응답은 유지한체 질의를 단락 내 다른 질의로 대체함으로써 부정 샘플을 생성할 수 있다.

[0045] 주어진 단락 $Psg$와 이에 대응하는 GT QA쌍 집합 $(Q^{\text{gt}}, A^{\text{gt}})$에 대하여, 긍정 샘플 집합 $POS = \{(q_i^{\text{gt}}, a_j^{\text{gt}}) \mid q_i^{\text{gt}} \in Q^{\text{gt}}, a_j^{\text{gt}} \in A^{\text{gt}}, i = j\}$과 부정 샘플 집합 $NEG = \{(q_i^{\text{gt}}, a_j^{\text{gt}}) \mid q_i^{\text{gt}} \in Q^{\text{gt}}, a_j^{\text{gt}} \in A^{\text{gt}}, i \neq j\}$을 구축할 수 있다. 실시예에 따르면, 다른 단락 내의 QA쌍을 용이한 부정 케이스로 고려하고, 이를 순위 생성기 학습에서는 부정 샘플로 포함시키지 않을 수 있다.

[0046] 그런 다음, QA쌍과 그에 대응하는 단락은, $\theta_{Rank}$를 학습하기 위한 입력 시퀀스(input sequences)를 구축하기 위해, 연결된다(concatenated). 이 입력 시퀀스를 피딩(feeding)함으로써, $\theta_{Rank}$는 부정(negative) 및 긍정(positive)을 나타내는 이진 레이블(binary labels)을 분류하도록 학습된다.

[0047] 추론 단계에서, $\theta_{Rank}$는 입력 QA쌍이 각각 긍정과 부정으로 분류될 스코어를 반환한다. 두 스코어를 참조함으로써, 각 QA쌍을 순위 매길 수 있다. 보다 구체적으로, 긍정 샘플과 부정 샘플 간 1/0(또는 0/1)로 이진 분류하도록, 순위 생성기가 학습된다. 추론 시에는 앞서 생성한 질의응답쌍에 대해 순위 생성기의 입력으로 단락과 함께 넣어주게 되며, 순위 생성기는 1/0 출력 직전의 마지막 은닉 상태(last hidden stat) 중 CLS 토큰에 해당하는 값을 추출하여 소프트맥스(softmax)를통과시킨 뒤 각 레이블로 예측될 확률 값을 출력할 수 있다. 1(또는 0)로 예측될 확률이 높은 순서대로 나열한 후에는 중복 제거(또는 중복 완화) 과정이 수행될 수 있다.

[0048] 이러한 프로세스를 통해, 순위 생성기는, QA쌍들 간의 높은 상관 관계(correlation)와 대응 단락에 대한 높은 관련성(relevance)을 나타내는 데이터 선택을 우선시하도록 학습된다.

[0049] **중복 제거(Overlap Mitigation)** 순위 생성 모델이 QA쌍들의 관련성을 향상시키는 동안, 상위 순위의 QA쌍들이 유사한 형태를 구성하는 중복 이슈(issue of duplication)가 존재한다. 이 이슈를 완화하기 위해, QA쌍 후보에서 응답의 어휘적 중복(lexical overlap)을 감소시키기 위해 재조정된 순위 점수(re-scaled ranking score)를 계산한다.

[0050] 순위 생성 모델을 사용하여 계산된 높은 점수의 순서대로 QA쌍을 순차적으로 선택한다. 각 선택 프로세스에서 어휘적 중복을 고려하기 위해, 선택한 쌍과 이전에 선택된 QA쌍들 사이의 Rouge-L 스코어를 측정한다. 순위 생성 모델에 의해 측정된 각 쌍의 점수 s는 s - Rouge-L * abs(s)로 재조정된다(re-scaled). 이러한 프로세스를 통해, 이전에 선택된 QA쌍들과 높은 어휘적 중복을 나타내는 QA쌍들의 점수를 다운스케일한다(down-scale). 이는 순위 생성 모델에 의해 계산된 스코어를 반영하면서도 다양한 유형의 QA를 선택할 수 있도록 한다. 중복 완화 알고리즘의 자세한 절차는 알고리즘 1(Algorithm 1)에 제시되어 있다.

[0051] 구체적인 예로, 1위 QA쌍(t1)과 나머지 QA쌍들 간의 Rougue-L 스코어가 측정된다. 이후, 재조정된 스코어를 기준으로 가장 높은 값을 갖는 QA쌍을 2위 QA쌍(t2)로 선택한다. t1 및 t2에 대해, 이들을 제외한 나머지 QA쌍들과의 Rougue-L 스코어를 측정하며, 측정된 Rougue-L 스코어로 재조정된 스코어를 산출한다. 물론, t1과의

Rougue-L 스코어 및 t1과의 Rougue-L 스코어를 이용한 스코어 재조정은 이미 수행되었으므로, 생략될 수도 있다. 다시, 재조정된 스코어가 가장 높은 QA쌍(t3)을 선택한다. 이와 같은 과정은, 최종 top-K 개의 QA쌍이 선택될 때까지 반복적으로 수행될 수 있다. 여기서, K는 매개변수이다.

[0052] [알고리즘 1]

**Algorithm 1** Overlapping based reranking

**Given**: Passage $Psg$,
**Input**: Generated QA pair $QA^{gen} = \{(q_i, a_i)\}_{i=1}^{N}$
**Parameter**: int k
**Define**: $\text{score}_i \leftarrow \text{RankingModule}(q_i, a_i, Psg)$
**Choose**: criterion $_i \leftarrow q_i$ or $a_i$
**Choose**: Metric ← ROUGE-L or BLEU

output ← [], comparing ← []
**while** len(output) ≤ k **do**
  **for** $(q_j, a_j)$ in $QA^{gen}$ **do**
    **if** comparing is not EMPTY **then**
      $\text{overlaps}_j$ = [Metric(criterion$_j$, item)
                for item in comparing]
      $\text{overlap}_j = \max(\text{overlaps}_j)$
      **Define**: $\text{score}_j^* \leftarrow \text{score}_j - \text{overlap}_j * |\text{score}_j|$
    **else**
      **Define**: $\text{score}_j^* \leftarrow \text{score}_j$
    **end if**
  **end for**
  $(q_i, a_i) \leftarrow$ **Pick from** $QA^{gen}$ with highest $\text{score}_j^*$
  output ← **Append** $(q_i, a_i)$
  comparing ← **Append** $\text{criterion}_i$
  $QA^{gen} \leftarrow$ **Pop** $(q_i, a_i)$
**end while**
**return** output

[0053]

[0054] 중복 완화의 세부적인 과정은 다음과 같다. 먼저 Criterion(기준)과 Metric(메트릭)을 정의한다. Criterion(기준)은 질의나 응답 중에서 중복 확인 대상이 되는 문장을 의미하며, Metric(메트릭)은 중복을 측정하는 평가 메트릭을 의미한다. Metric(메트릭)으로 Rouge-L 또는 BLEU를 사용하는 것을 제안한다. 주요 실험에서, $\text{criterion}_i$를 $a_i$로(즉, QA쌍의 응답) 선택하고 Metric(메트릭)으로 ROUGE-L을 선택한다. 이 과정에서 Metric(메트릭)은 0에서 1 사이의 중복 점수(overlap score)를 반환한다. 중복을 평가할 때, 모든 문장을 레마타이즈(lemmatize)하고, 모든 QA쌍에서 모든 불용어(stop words)를 제거한다.

[0055] 도 2는 본 발명의 일 실시예에 따른 질의응답쌍 생성 방법을 설명하기 위한 흐름도이다.

[0056] 도 2을 참조하면, 질의응답쌍 생성 방법은 적어도 프로세서(processor) 및/또는 메모리(memory)를 포함하는 컴퓨팅 장치에 의해 수행될 수 있다. 따라서, 질의응답 쌍 생성 방법을 이루는 단계들 중 적어도 일부는 질의응답쌍 생성 장치로 명명될 수도 있는 컴퓨팅 장치에 포함되는 프로세서의 동작으로 이해될 수도 있다. 컴퓨팅 장치는 PC(Personal Computer), 서버(server), 랩탑 컴퓨터, 태블릿 PC 등을 포함할 수 있다. 컴퓨팅 장치는 물리적으로 분리되는 복수의 장치들로 수행되거나 클라우드 상에서 제공될 수도 있다. 이하에서, 질의응답쌍 생성 방법을 설명함에 있어 앞선 기재와 중복되는 내용에 관하여는 그 상세한 기재를 생략하기로 한다.

[0057] 우선, 주어진 지문(passage)에 대한 질의집중요약이 생성된다(S110). 질의집중요약은 지문에 포함된 문장의 개수만큼 생성될 수 있다. 즉, 질의집중요약 모델은 지문과 문장을 입력받고, 문장에 대응하는 질의집중요약을 출력할 수 있다.

[0058] 지문과 질의집중요약을 기초로, 초기 정답(또는 초기 응답)이 생성된다(S120). 즉, 정답 생성기 또는 정답 생성 모델은 지문과 질의집중요약을 입력받고 이에 대응하는 초기 정답을 출력할 수 있다. 초기 정답은 질의집중요약에 대응되기 때문에, 질의집중요약의 개수에 해당하는 초기 정답이 생성된다.

[0059] 초기 정답에 대응하는 질의가 생성된다(S130). 질의는 미리 정해진 의문사를 포함하도록 생성되기 때문에, 하나의 초기 정답에 대하여 복수개(예컨대, 6개)의 질의가 생성될 수 있다. 즉, 질의 생성기 또는 질의 생성 모델은 초기 정답, 지문, 및 의문사를 입력받고, 이에 대응하는 질의를 출력할 수 있다.

[0060] 다음으로, 생성된 각 질의에 대응하는 응답이 생성된다(S140). 응답은 입력되는 질의에 대응되기 때문에, 질의

의 개수만큼의 응답이 생성된다. 즉, 질의응답 모델은 질의와 지문을 입력받고 질의에 대응하는 응답을 출력한다.

[0061] 생성된 질의응답쌍들 중 최종 질의응답쌍이 선택된다(S150). 질의응답쌍의 선택 과정에 대한 상세한 내용은 상술하였기 때문에, 이에 대한 설명은 생략하기로 한다.

[0062] 이하에서는, 실험(Experiments)을 상세히 설명한다.

[0063] **실험 설정**

[0064] **Dataset** 실험에서, FairytaleQA 데이터셋(Xu et al., 2022)을 활용한다. FairytaleQA는 어린이 스토리북(storybook) 학습 및 평가를 위해 특별히 설계된 데이터셋으로, 본 발명의 교육 목적과 일치한다. 데이터 구축 과정에서, 교육 전문가들이 QA쌍을 수동으로 작성하여 신뢰성(reliability)과 타당성(validity)을 보장하였다. 훈련, 검증 및 테스트 세트는 각각 232 권의 책에서 유래된 8,548개의 QA쌍, 23 권의 책에서 유래된 1,025개의 QA쌍, 및 23 권의 책에서 유래된 1,007개의 QA쌍을 포함한다. 데이터셋에서 제시된 서술적인 요소(narrative elements) (예: 캐릭터(character), 설정(setting), 행동(action) 등) 대신, 의문문을 기반으로 질문을 다양화하여 이러한 요소 외에도 확장된 유형의 질문을 유도한다. 기존의 응답 유형은 서로 배타적이기 때문에 그대로 사용한다.

[0065] **Models** 제안하는 프레임워크를 구성하는 모든 모델은 FairytaleQA 데이터셋으로 학습된다. 질의집중요약 모델(QFS 모델)의 경우, Vig et al. (2021)에서 제공하는 모델 체크포인트(model checkpoints)를 사용하여 요약을 생성한다. AGM, QGM, 및 QAM의 학습에서, 사전학습된(pre-trained) BART-large(Lewis et al., 2020) 모델과 Fairseq에서 제공하는 프레임워크를 활용한다. 하이퍼파라미터의 경우, 2048 최대 토큰(max tokens), 조기 중지(early stopping) 10, 및 다항식 감쇠 스케줄러(polynomial decay scheduler)가 채택된다. 학습률(learning rate)과 드롭아웃(dropout)은 각각 AGM과 QGM에서 3e-05와 0.1, QAM에서 2e-05와 0.2로 설정한다. 모든 모델은 2개의 RTX8000 GPU에서 학습된다. 순위 생성 모델에는, RoBERTa-base(Liu et al., 2019) 모델과 Huggingface 프레임워크를 사용한다. 5회 에포크 동안 고정된 학습률 5e-07로 랭커를 훈련하며, 훈련에는 단일 GPU가 사용된다.

[0066] **평가 지표(Evaluation Metrics)**

[0067] 평가 메트릭으로는, Yao et al. (2022)가 사용한 주요 메트릭으로 MAP@N 스코어를 채택한다. Rouge-L을 사용한 MAP@N은 각 GT QA쌍과 상위 N개의 생성된 QA쌍 간의 Rouge-L을 계산하여 추가된 최대 점수 세트의 평균값을 나타낸다. 이 과정에서 QA쌍 내의 각 질의와 응답은 연결된다(concatenated). 그러나, MAP@N이 Yao et al. (2022)에서처럼 Rouge-L 정밀도 점수(prcision score)에 의해 측정될 때, 짧은 결과가 유리하다. 이는, 이 메트릭이 후보(candidates)의 수를 기준으로 최장 중첩(longest overlap)을 측정하기 때문이다. 정밀도 대신에, 정확한 측정을 위해 F1 점수를 선택한다. N-gram 중첩을 기반으로 한 메트릭은 품질을 보장하지 않기 때문에, 의미론적 동등성(semantic equivalence)을 평가하기 위해 유사성 점수(similarity scores)에 기반한 BERTScore를 MAP@N에 추가적으로 채택한다.

[0068] **Baselines**

[0069] 두 개의 교육용 QAG 시스템을 베이스라인으로 채택한다.

[0070] **FQAG** FQAG (Yao et al., 2022)는 FairytaleQA의 최신 연구 중 하나로, 응답 생성, 질의 생성, 및 랭킹 모듈을 포함하는 세 단계 파이프라인을 통해 QAG를 수행한다. 재구현을 위해, 제공된 체크포인트를 로드하여 FairytaleQA의 검증 및 테스트 세트에 대한 QA쌍을 생성한다.

[0071] **SQG** SQG(Dugan et al., 2022)은 교육용 QAG에서 최근에 발표된 논문으로, 주어진 단락의 요약을 활용한다. QA쌍은 응답 생성, 질의 생성, 및 질의 응답 모델을 활용하여 생성된다. 이 경우, 상위 N의 수를 맞추기 위해, 생성된 순서에 따라 QA쌍을 선택하거나 빔 크기를 조정하여 출력을 늘린다.

[0072] 이하에서는, 실험 결과 및 분석(Results and Analysis)을 설명한다.

[0073] **자동 평가(Automated Evaluation**

[0074] **Result on MAP@N with Rouge-L** 표 1은 QAG 시스템에 따른 Rouge-L F1 점수와 함께 MAP@N의 주요 결과를 보여준다. 결과적으로, 본 발명에서 제안하는 시스템은 모든 분할(splits)과 상위 N 결과에서 기준 모델(baseline model)을 획기적으로 능가한다. 특히 테스트 세트에서, 상위 10에서 +0.068, MAP@5에서 +0.055, 그리고 MAP@3

에서 +0.048로, FQAG를 크게 앞지르고 있다. SQG는 FQAG보다 우수한 결과를 달성하지만, 여전히 제안 시스템을 능가하지 못한다. SQG와 비교하면, 제안 시스템은 주로 0.455에서 0.503으로 모든 상위 N 결과에서 개선됨을 보여준다(+0.048). 이러한 결과는, 다양한 QA쌍 후보를 생성하고 타당한 쌍을 적절하게 설정하는 것이 성능 향상에 기여하는 요소 중 하나로 작용한다는 것을 시사한다.

[0075] [표 1]

[0076]

| | MAP@N (Rouge-L F1) | | | | MAP@N (BERTScore F1) | | | |
|---|---|---|---|---|---|---|---|---|
| Method | Top 10 | Top 5 | Top 3 | Top 1 | Top 10 | Top 5 | Top 3 | Top 1 |
| *FQAG*(Yao et al., 2022) | 0.440 / 0.435 | 0.375 / 0.374 | 0.333 / 0.324 | 0.238 / 0.228 | 0.9077 / 0.9077 | 0.8990 / 0.8997 | 0.8929 / 0.8922 | **0.8768** / 0.8776 |
| *SQG*(Dugan et al., 2022) | 0.460 / 0.455 | 0.392 / 0.388 | 0.344 / 0.337 | 0.234 / 0.242 | 0.9056 / 0.9062 | 0.8953 / 0.8955 | 0.8876 / 0.8878 | 0.8707 / 0.8723 |
| Ours | **0.500 / 0.503** | **0.426 / 0.429** | **0.369 / 0.372** | **0.247/ 0.254** | **0.9156 / 0.9178** | **0.9046 / 0.9068** | **0.8956 / 0.8977** | 0.8752 / **0.8783** |

[0077] **Result on MAP@N with BERTScore** BERTScore를 사용하여 GT와 생성된 QA쌍 간의 의미적 동등성(semantic equivalence)을 평가하여 MAP@N을 측정한다. 즉, MAP@N을 측정할 때 Rouge-L F1 점수 대신 BERTScore의 F1 값을 사용한다. 결과적으로, 본 발명에서 제안하는 시스템은 MAP@1 검증 결과를 제외한 모든 설정에서 더 높은 성능을 달성한다. 테스트 결과 중에서 가장 우수한 경우인 MAP@10에서 FQAG와 SQG는 각각 0.9077과 0.9062를 보여주었고, 제안 기법은 0.9178을 기록하여 +0.0101 및 +0.0116으로 능가하였다. 제안 기법의 성능이 가장 높은 경향은 Rouge-L F1 결과와 일관되어 있다. 그러나, FQAG가 SQG보다 BERTScore에서 더 높은 성능을 보고하는 것을 관찰할 수 있다. 성능 차이는 미세하지만, 이 결과는 FQAG의 생성된 QA쌍이 SQG보다 의미적으로 우수하다는 것을 시사한다.

[0078] **통계적 평가(Statistical Evaluation)**

[0079] 생성된 QA쌍의 질의 및 응답 유형 다양성을 평가하기 위해, 통계적 평가를 수행한다. 결과적으로, 본 발명에서 보고된 질의 유형은 다른 모델보다 더 균형있는 편이다. 일반적으로 'what'과 'who'를 생성하는 다른 모델과 달리, 제안하는 QAG 시스템은 추론을 필요로 하는 'why'와 'how'와 같은 다양한 유형의 질문으로 균형 잡혀 있다. 이는, 어린이들이 다양한 관점에서 생각하는 데 있어 다양한 유형의 질의를 받아볼 수 있는 잠재력을 시사한다.

[0080] 응답 유형에 대해서는, 제안 시스템이 32.06%의 암시적 응답(implicit answers)을 포함하고 있으며, 이는 암시적 답변이 잘 생성되었음을 나타낸다. 이는, 모델이 어린이들의 평가를 균형있게 도울 수 있게 한다. 반면에 다른 모델들은 응답 범위 추출 방법을 사용하며, 결과적으로 암시적 답변이 0%이다.

[0081] **인간 평가(Human Evaluation)**

[0082] 더 자세한 검토를 위해, 인간 평가를 수행한다. 각 단락마다, 학사 학위 또는 교육 전문가 자격을 가진 세 명의 인간 평가자가 GT와 세 가지 QAG 시스템에서 생성된 각각의 세 개의 QA쌍을 평가한다. 총 20개의 단락에 대한 인간 평가를 수행하며, GT와 SQG에 대해 순차적으로 세 개의 QA쌍을 선택한다. 다음은 인간 평가에 사용되는 기준(criteria)이다. 전역 설정(global setting)에서 평가자에게 전체 시스템을 순위 지정하도록 지시하고, 지역 케이스(local cases)에서는 각 시스템에서 생성된 세 개의 QA쌍 중 몇 개가 속성 항목과 일치하는지 선택하도록 한다.

[0083] (전역 설정) **다양성-Q(Diversity-Q):** 질의 다양성 측면에서 GT와 세 개의 QAG 시스템의 생성 결과를 순위 지정한다. **다양성-A(Diversity-A):** 응답 다양성 측면에서 GT와 세 개의 QAG 시스템의 생성 결과를 순위 지정한다. **품질-E(Quality-E):** 전반적인 관점에서 전체 시스템의 품질을 순위 지정한다.

[0084] (지역 설정) **적절성(Relevancy):** 단락과 QA쌍 간의 관련성을 평가한다. 질의 또는 응답 중 하나라도 관련이 없으면 관련성이 없는 것으로 간주된다. **수용성(Acceptability):** 생성된 질의과 해당 응답이 올바르게 생성되었는지를 평가한다. 지문과의 관련성은 고려되지 않으며, 둘 중 하나라도 어색하면 올바르지 않은 것으로 간주된다. **유용성(Usability):** 생성된 QA쌍이 교육 목적으로 사용될 수 있는지를 평가한다. **가독성(Readability):** 생성된 QA쌍이 문법적으로 정확한지를 평가한다. **난이도(Difficulty):** 생성된 QA쌍이 과도하게 쉬운지를 평가한다.

[0085] 표 2는 인간 평가 결과를 제시한다. 본 발명에 의한 접근 방식은 질의 및 응답 다양성 측면에서 높은 성능을 달성하며, 각각 평균 순위가 2.35와 2.18이다. 전역 설정에서, FQAG의 Quality-E는 2.66이고 SQG는 3.30이지만, 제안 시스템은 2.35의 점수로 그들을 능가한다. 이러한 결과는 제안하는 QAG가 다양성을 향상시키면서 다른 시스템과의 직접적인 비교에서 양적으로나 질적으로도 뛰어나다는 것을 보여준다. 지역 설정의 결과는 제안 기법이 가독성을 제외하고 FQAG와 SQG를 능가한다는 것을 나타낸다. 생성된 QA쌍의 평가 결과로서 지문과 생성된 QA 간의 관련성(2.69), 질의와 응답의 수용성(2.22), 및 교육 목적의 유용성(1.9)은 다른 시스템과 비교했을 때 가

장 높은 결과를 보여준다. 난이도의 경우에는, GT보다 약간의 성능 향상을 관찰할 수 있다. 그러나, 가독성에서는 제안 기법의 결과가 기존 모델의 2.64와 2.55보다 낮은 2.35를 보여준다. 제안 기법으로 생성한 QA쌍의 평균 길이가 더 길어서 난이도와의 작은 트레이드 오프가 발생한 것으로 추측된다. 결과적으로, 생성된 QA쌍이 품질뿐만 아니라 다양성을 확보하는 데 실제로 효과적이라고 결론지을 수 있다.

[0086] [표 2]

| | *global* | | | *local* | | | | |
|---|---|---|---|---|---|---|---|---|
| **Method** | **Diversity-Q ↓** | **Diversity-A ↓** | **Quality-E↓** | **Relevancy ↑** | **Acceptability ↑** | **Usability ↑** | **Readability ↑** | **Difficulty ↑** |
| *FQAG*(Yao et al., 2022) | 3.03 | 3.06 | 2.66 | 2.65 | 2.14 | 1.74 | **2.64** | 1.11 |
| *SQG*(Dugan et al., 2022) | 2.96 | 3.03 | 3.30 | 2.44 | 1.87 | 1.34 | 2.55 | 1.36 |
| Ours | **2.35** | **2.18** | **2.35** | **2.69** | **2.22** | **1.9** | 2.35 | **1.98** |
| GT | 1.65 | 1.71 | 1.68 | 2.97 | 2.65 | 2.50 | 2.80 | 1.95 |

[0087]

[0088] 이상에서 설명된 장치는 하드웨어 구성 요소, 소프트웨어 구성 요소, 및/또는 하드웨어 구성 요소 및 소프트웨어 구성 요소의 집합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치 및 구성 요소는, 예를 들어, 프로세서, 콘트롤러, ALU(Arithmetic Logic Unit), 디지털 신호 프로세서(Digital Signal Processor), 마이크로컴퓨터, FPA(Field Programmable array), PLU(Programmable Logic Unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 하나 이상의 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(Operation System, OS) 및 상기 운영 체제 상에서 수행되는 하나 이상의 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술 분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(Processing Element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 콘트롤러를 포함할 수 있다. 또한, 병렬 프로세서(Parallel Processor)와 같은, 다른 처리 구성(Processing Configuration)도 가능하다.

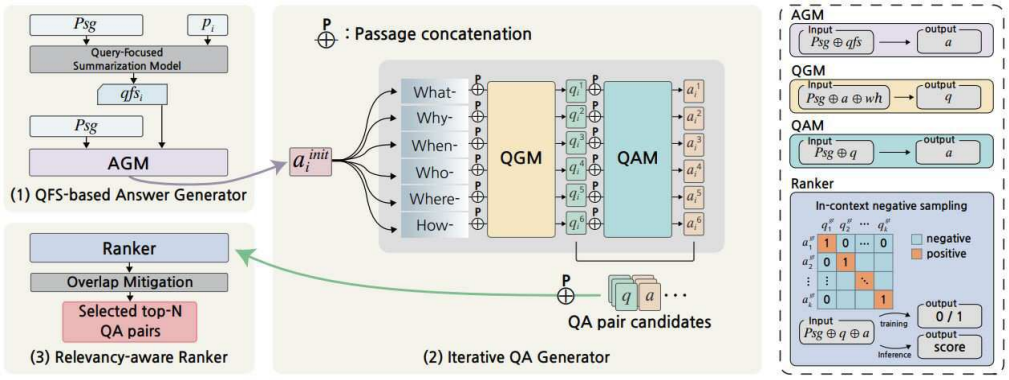
[0089] 소프트웨어는 컴퓨터 프로그램(Computer Program), 코드(Code), 명령(Instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(Collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성 요소(Component), 물리적 장치, 가상 장치(Virtual Equipment), 컴퓨터 저장 매체 또는 장치, 또는 전송되는 신호 파(Signal Wave)에 영구적으로, 또는 일시적으로 구체화(Embody)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 이상의 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.

[0090] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(Magnetic Media), CD-ROM, DVD와 같은 광기록 매체(Optical Media), 플롭티컬 디스크(Floptical Disk)와 같은 자기-광 매체(Magneto-optical Media), 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다. 상기된 하드웨어 장치는 실시예의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

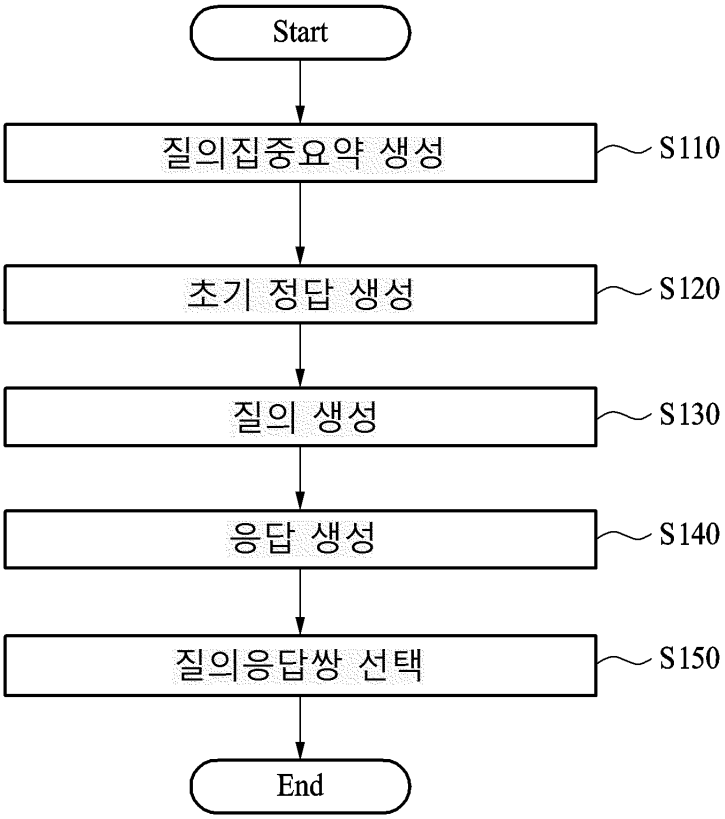
[0091] 본 발명은 도면에 도시된 실시예를 참고로 설명되었으나 이는 예시적인 것에 불과하며, 본 기술 분야의 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시예가 가능하다는 점을 이해할 것이다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성 요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성 요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다. 따라서, 본 발명의 진정한 기술적 보호 범위는 첨부된 등록청구범위의 기술적 사상에 의해 정해져야 할 것이다.

## 도면

### 도면1


Psg
Query-Focused Summarization Model
AGM
(1) QFS-based Answer Generator
Ranker
Overlap Mitigation
Selected top-N QA pairs
(3) Relevancy-aware Ranker
: Passage concatenation
What-
Why-
When-
Who-
Where-
How-
QGM
QAM
QA pair candidates
(2) Iterative QA Generator
AGM
QGM
QAM
Ranker
In-context negative sampling
negative
positive
training
Inference
score

### 도면2


Start
질의집중요약 생성
S110
초기 정답 생성
S120
질의 생성
S130
응답 생성
S140
질의응답쌍 선택
S150
End