

---

# 때때로 연골이 아파오지만

S(tanislav)\_RAM

---

# INDEX

---

- AI 모델 튜닝
    - 학습 데이터 선정 및 DB 수집
    - 데이터 필터링
    - 파인 튜닝
  - 작품 제작
    - 작품 개요
    - 작품 작성 방법
-

---

# AI 모델 튜닝

---

# 학습 데이터 선정 및 수집

## 데이터 선정 기준

1. 문장의 수준이 높은가?
2. 데이터의 fit이 잘 맞는가?
3. 데이터의 양이 많은가?

### 문장 웹진 [소설]

등단 소설가들의 청탁 소설로 문장 수준 높음  
에세이가 아닌 소설이란 점에서 fit은 보통  
데이터 양은 약 단편소설 500편으로 매우 많음

### 브런치 [에세이]

최소한의 검증을 거친 작가진으로 문장 수준 보통  
에세이류가 많으므로 fit은 높음  
데이터 양이 매우 많음

### 거울 웹진 [소설]

장르 소설가들의 단편 소설 위주로 문장 수준 보통  
소설이란 점에서 fit은 보통  
출간 작가의 데이터로 한정할 경우 양이 보통임.

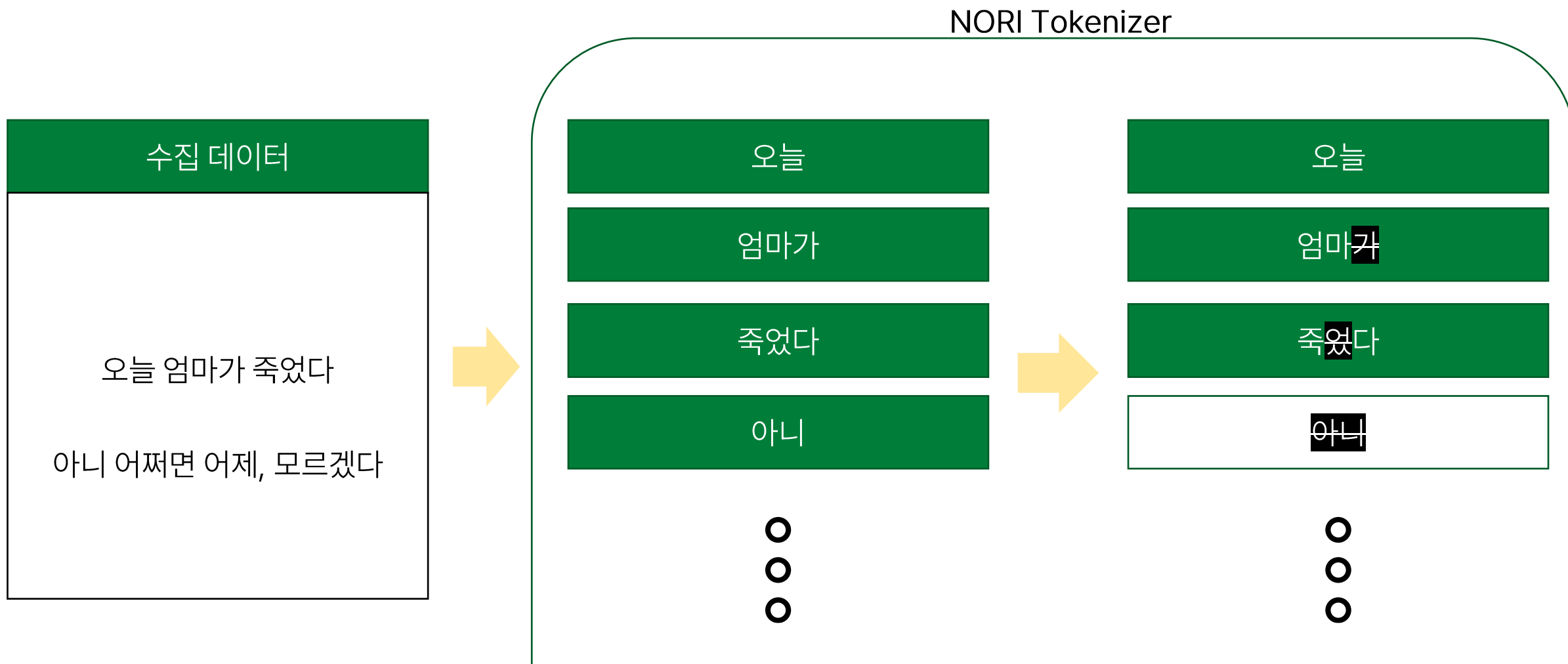
# 데이터 필터링 : 전처리

## KLUE

### Korean Language Understanding Evaluation

Korean Language Understanding Evaluation (KLUE) benchmark is a series of datasets to evaluate natural language understanding capability of Korean language models. KLUE consists of 8 diverse and representative tasks, which are accessible to anyone without any restrictions. With ethical considerations in mind, we deliberately design annotation guidelines to obtain unambiguous annotations for all datasets. Furthermore, we build an evaluation system and carefully choose evaluations metrics for every task, thus establishing fair comparison across Korean language models.

# 데이터 필터링 : Save

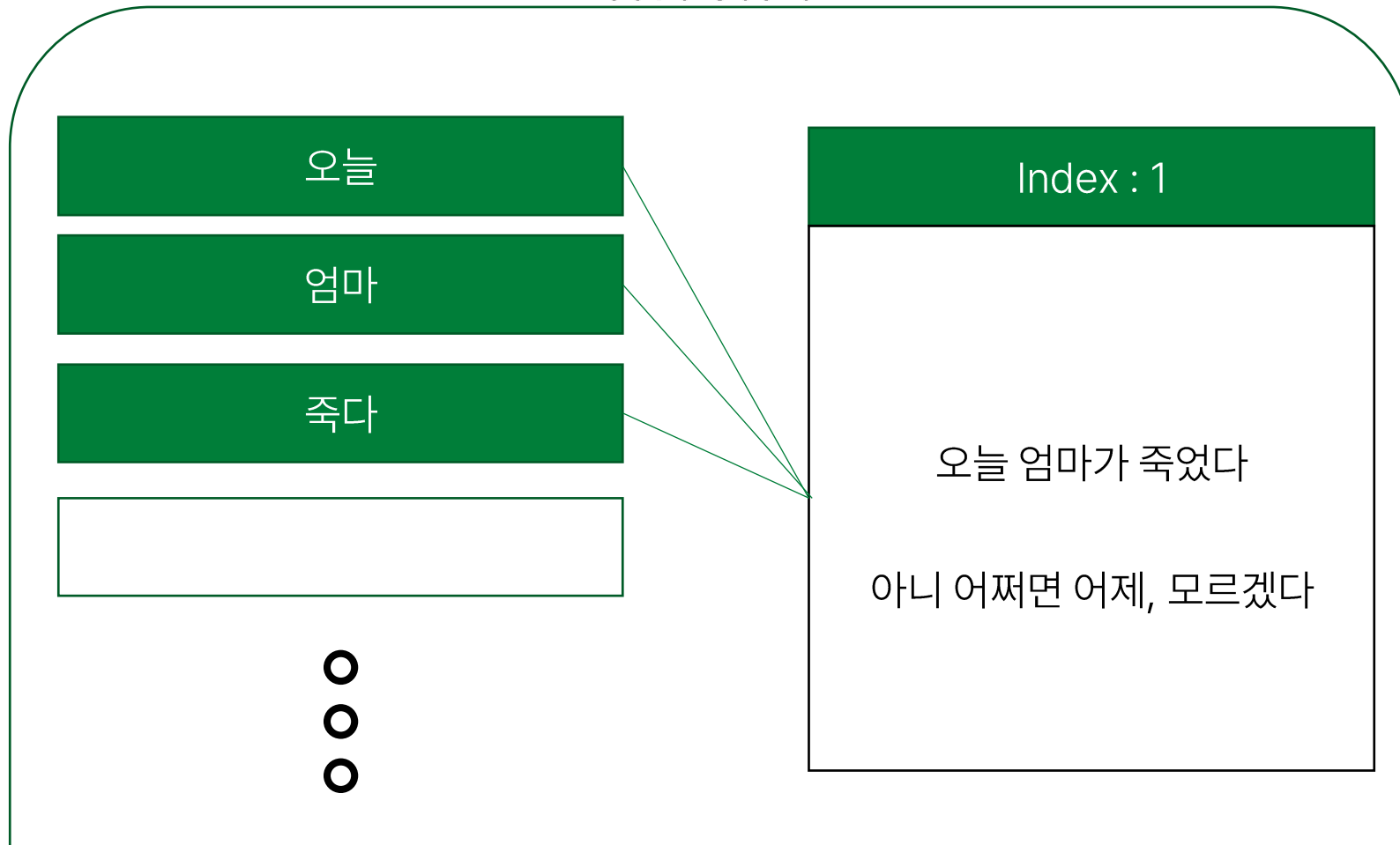


# 데이터 필터링 : Save

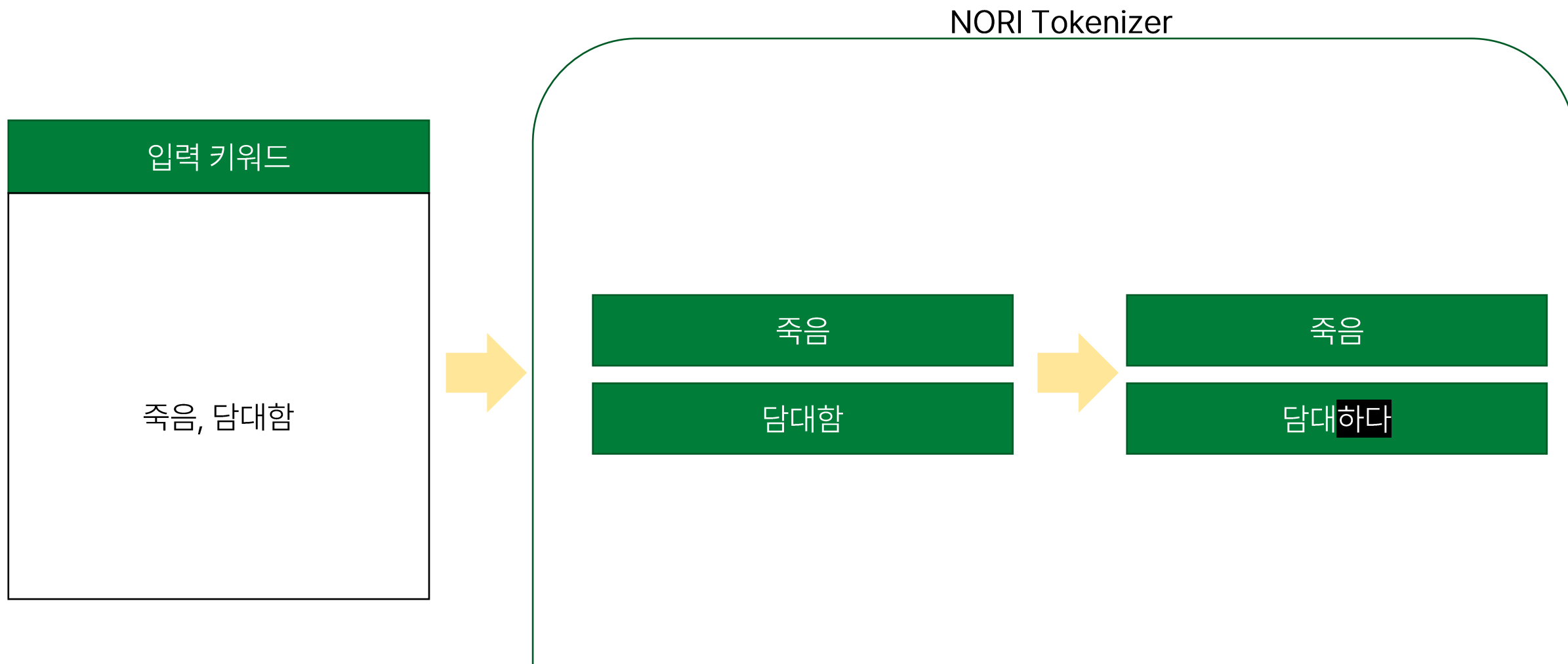
NORI Tokenizer



Elastic Search



# 데이터 필터링 : Query





# 데이터 필터링 : Query

NORI Tokenizer

죽음

담대하다

Elastic Search

Index : 1

죽음

담대하다

오늘 엄마가 죽었다

아니 어쩌면 어제, 모르겠다

○  
○  
○

# 파인 튜닝

## 정량적 파인 튜닝

- Loss와 accuracy가 빠르게 향상되는 하이퍼 파라미터 값을 찾아 학습을 수행
- KLUEBenchmark 태스크 Relation Extraction을 이용, F1 score 가 높아지는 방향 (텍스트에서 단어들 간의 의미론적 관계가 잘 식별되게 하는 방향)으로 학습 수행

## 정성적 파인 튜닝

- 생성된 글을 보고, 인간의 기준으로 생성이 잘 되었는지 판단
- 글이 잘 나오는 하이퍼 파라미터값을 찾아 반복 수행

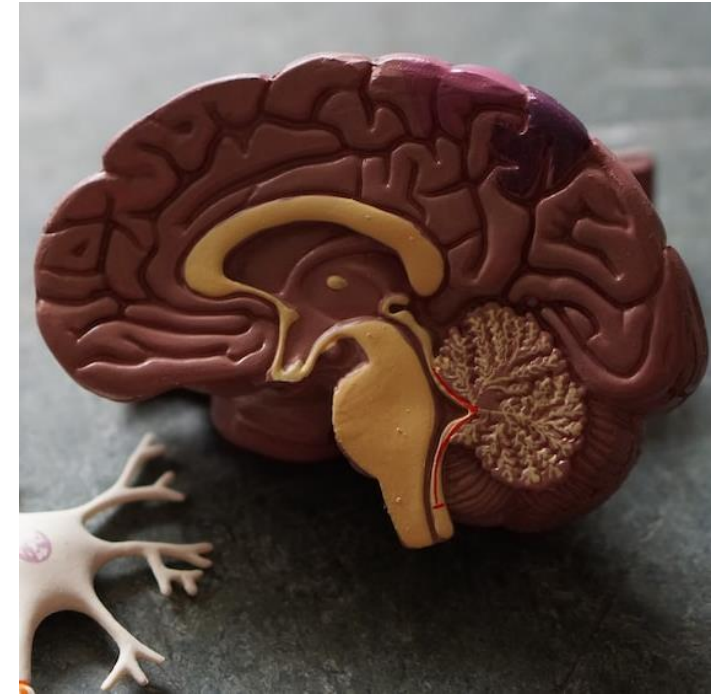
좋은 모델

---

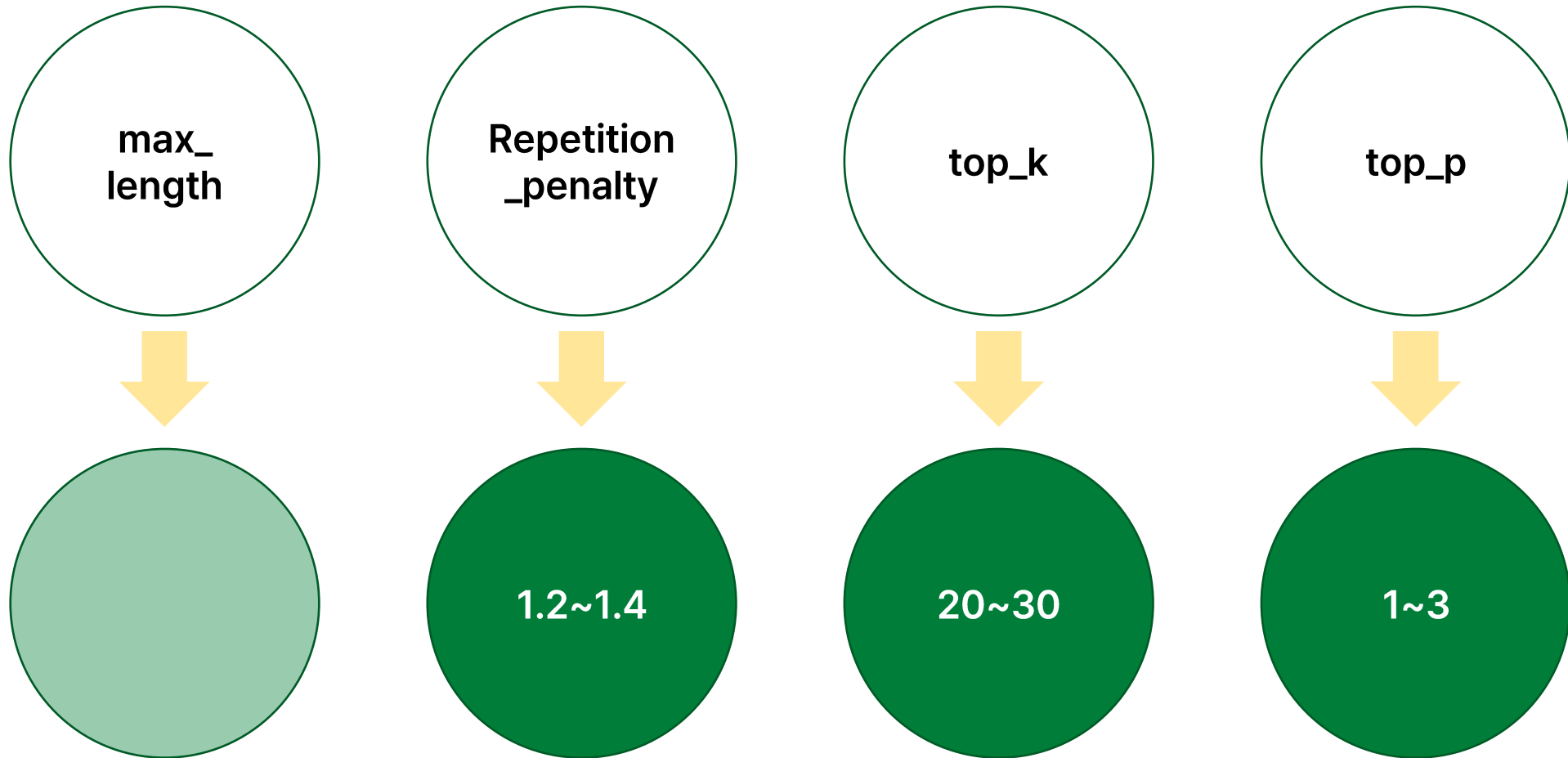
작품 제작

---

# 작품 개요

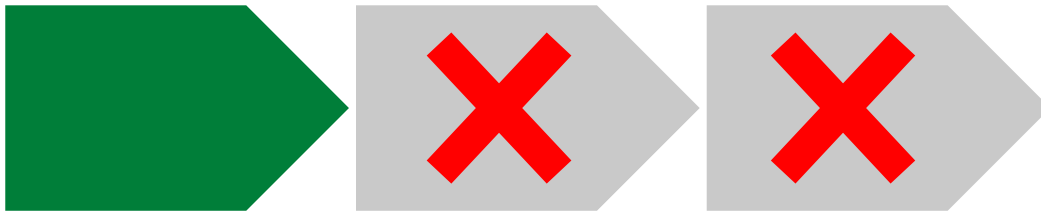


# 작품 작성 방법

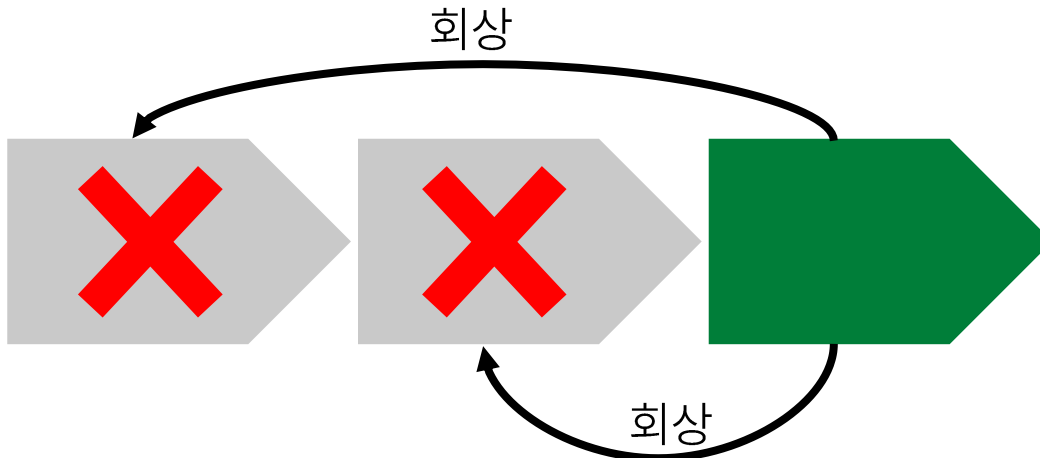


# 작품 작성 방법

사건의 발생  
서술



반복서술이나  
사유



- 온전히 생각만 하는 서술자, 사유가 온전치 못한 서술자를 내세워 혼란스러운 과거를 정당화
- 파편화된 회상이 문제없이 받아들여질 수 있도록 소위 '산책소설'적 방식으로 서술

---

질의응답

---