

검색해조 최종발표

- ★ 정새롬
- ★ 성재훈
- ★ 이태훈
- ★ 박제윤(멘토)

2022.10.24.

Presentation INDEX |

01 결과물 소개

02 작업 과정

03 성능 평가

04 추후 진행해야 할 사항

Presentation INDEX |

01 결과물 소개

02 작업 과정

03 성능 평가

04 추후 진행해야 할 사항

결과물 소개

Meilisearch 검색 결과 비교

개선된 Meilisearch 요약

검색결과 비교

신경정신 - Origin



Hits ~17

NAME	임상의를 위한 멘탈한방 입문
설명	일차진료 현장에서 신경정신 질환을 다룰 수 있도록 안내하는 책
NAME	대학경락경혈학 각론(상하권)
설명	3.남성생식계병증, 4.비뇨기계병증, 5. 신경정신 과병증,

신경정신 - Ours



Hits ~26

NAME	[보수교육]2022년 대한한방 신경정신 과학회 춘계 학술대회
설명	대한한방 신경정신 과학회 춘계
NAME	한의학정신건강센터 2021 월례회VOD
설명	경혈을 두드려 심리적 (...) 한방 신경정신 과

검색결과 비교

신경정신



Hits ~17

NAME	임상의를 위한 멘탈한방 입문
설명	일차진료 현장에서 신경정신 질환을 다룰 수 있도록 안내하는 책
NAME	대학경락경혈학 각론(상하권)
설명	3.남성생식계병증, 4.비뇨기계병증, 5. 신경정신 과병증,

Meilisearch - origin



검색어 인식 방법

-> 띄어쓰기 기반

“신경정신”

신경정신 분야에서

신경정신과

대한한방**신경정신과** ← 인식못함

한방**신경정신과** ← 인식못함

검색결과 비교

신경정신



Hits ~26

NAME	[보수교육]2022년 대한한방신경정신과학회 춘계 학술대회
설명	대한한방신경정신과학회 춘계
NAME	한의학정신건강센터 2021 월례회VOD
설명	경혈을 두드려 심리적 (...) 한방신경정신과

Meilisearch - Ours



검색어 인식 방법

-> 띄어쓰기 기반
+ 형태소 분석(사용자 사전)

“신경정신”

신경정신 분야에서

신경정신과

대한한방신경정신과

한방신경정신과

결과물 소개

Meilisearch 검색 결과 비교

개선된 Meilisearch 요약

개선된 Meilisearch 요약

검색방식

- 형태소 분석
- 사용자 사전

Mecab : 문서의 형태소분석 -> 검색의 정확성을 높임

사전 X : [('발', 'NNG'), ('뮈', 'NNG'), ('다', 'VCP+EC')]

사전 O : [('발뮈다', 'NNP')]

Settings

- 검색필드 한정
- 불용어
- 동의어
- 우선순위
- 오타허용

“Name” , “Description”

X

X

Default

Default

Presentation INDEX |

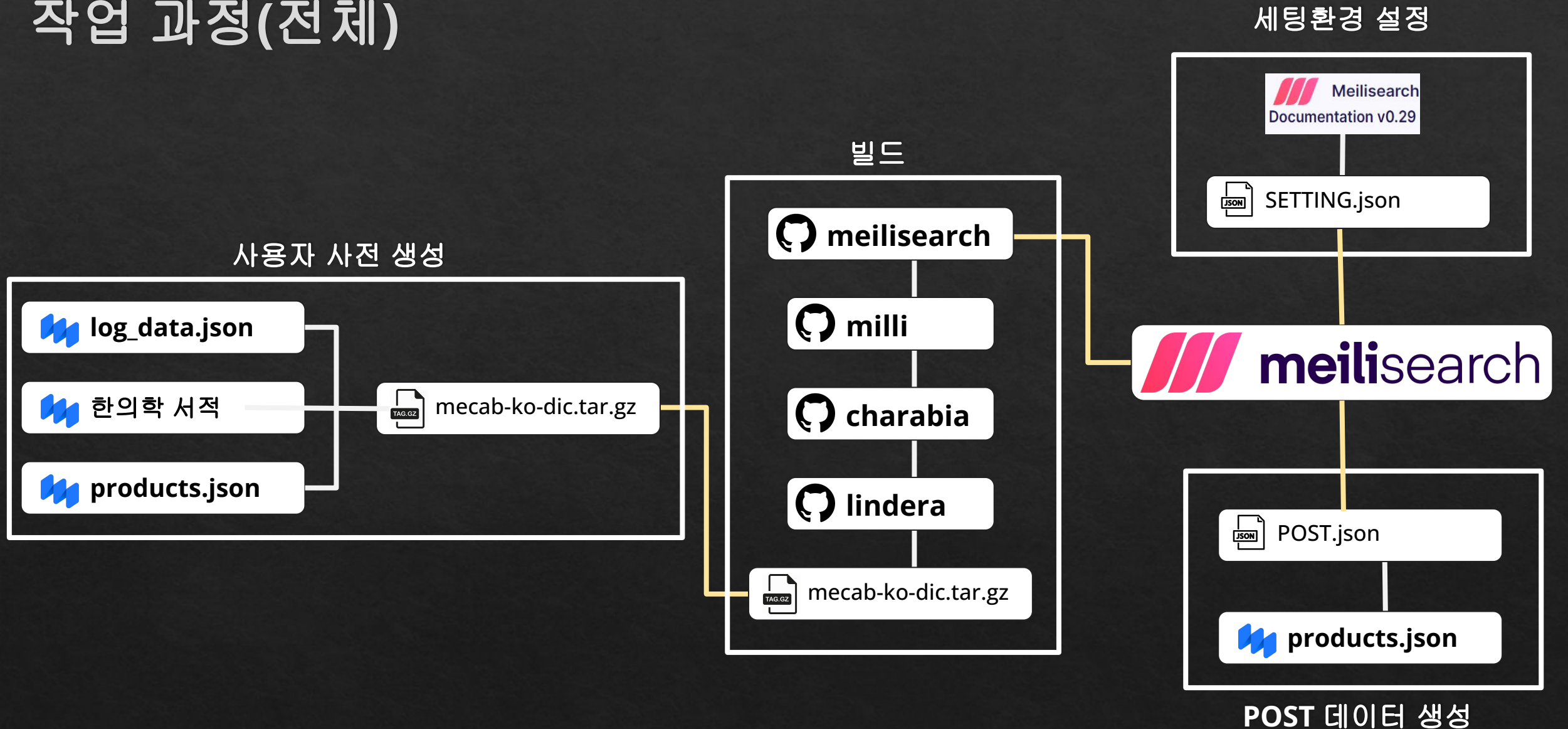
01 결과물 소개

02 작업 과정

03 성능 평가

04 추후 진행해야 할 사항

작업 과정(전체)



작업 과정

사용자 사전 생성

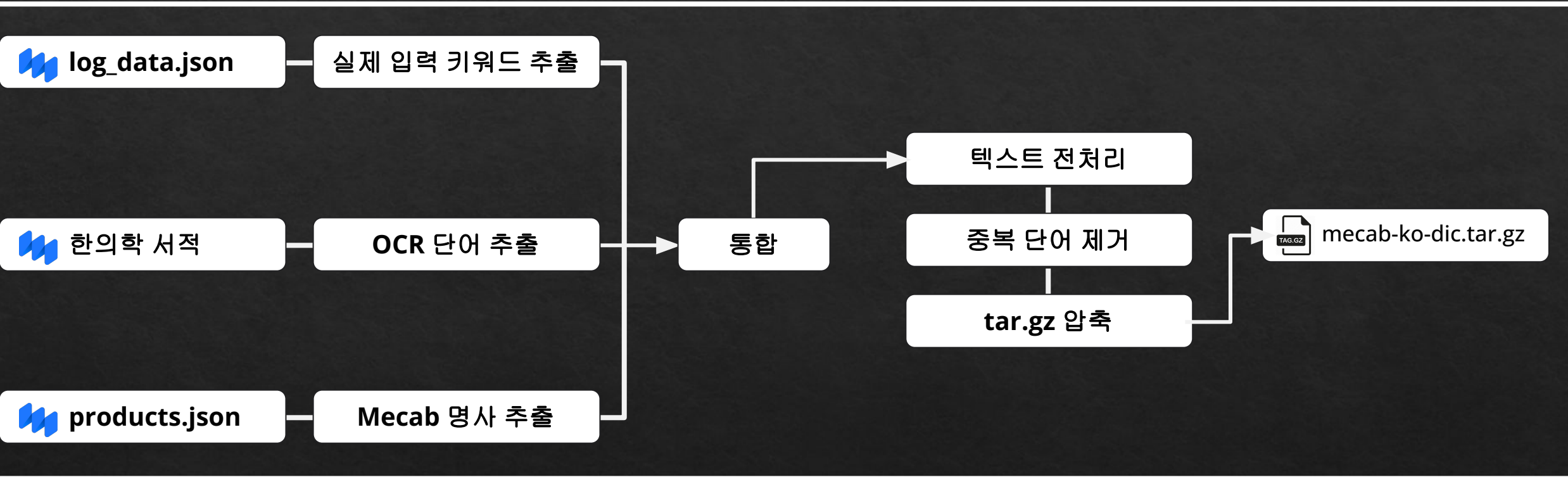
빌드

세팅환경

POST 데이터

빌드 과정(사용자 사전 생성)

사용자 사전 생성



작업 과정

사용자 사전 생성

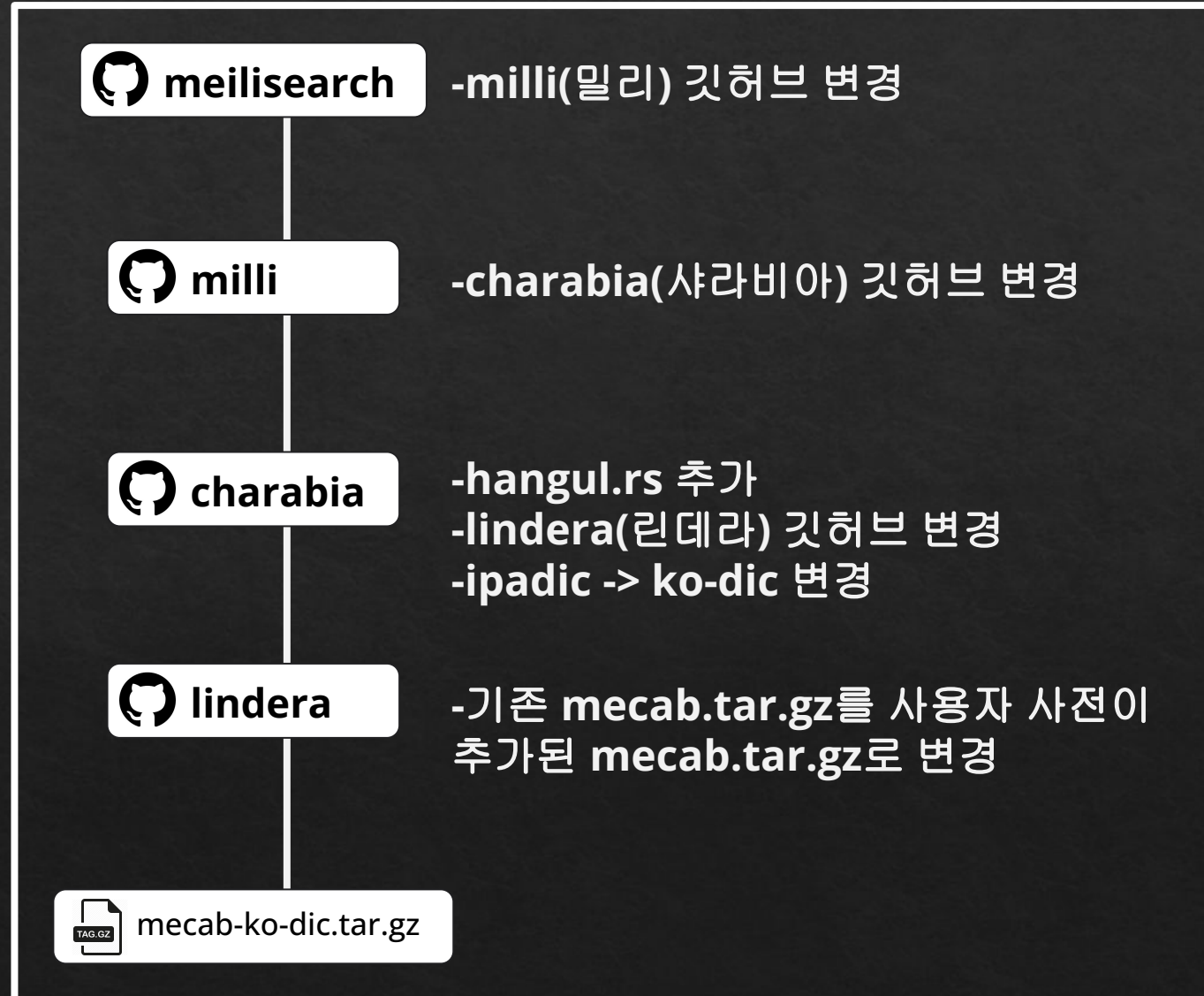
빌드

세팅환경

POST 데이터

빌드 과정(빌드)

빌드



작업 과정

사용자 사전 생성

빌드

세팅환경

POST 데이터

세팅환경 설정

세팅환경 설정



검색 필드 설정



불용어 설정



동의어 설정



우선순위 설정



오타허용 설정

세팅환경 설정 (검색 필드)



검색 필드 설정

검색 가능한 필드(컬럼) 수 : 37개

```
'_id', 'date_created', 'date_updated', 'images',  
'dimensions', 'name',  
  'description', 'meta_description',  
'meta_title', 'tags', 'attributes',  
  'enabled', 'discontinued', 'slug', 'sku',  
'code', 'tax_class',  
  'related_product_ids', 'prices', 'cost_price',  
'regular_price',  
  'sale_price', 'quantity_inc', 'quantity_min',  
'weight',  
  'stock_quantity', 'position',  
'date_stock_expected', 'date_sale_from',  
  'date_sale_to', 'stock_tracking',  
'stock_preorder', 'stock_backorder',  
  'category_id', 'category_ids', 'options',  
'variants'
```

검색할 수 있는 필드가 많으면 좋을까?

세팅환경 설정 (검색 필드)

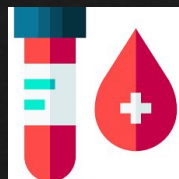


검색 필드 설정

Q 혈액검사 × data 5,141

	NAME	다나 일회용 부항컵 5pcs
	DESCRIPTION	
	_ID	615d2f664cf9a5002287de4c
	DATE_CREATED	2021-10-06T05:08:54.163Z
	DATE_UPDATED	2022-07-19T05:19:21.785Z

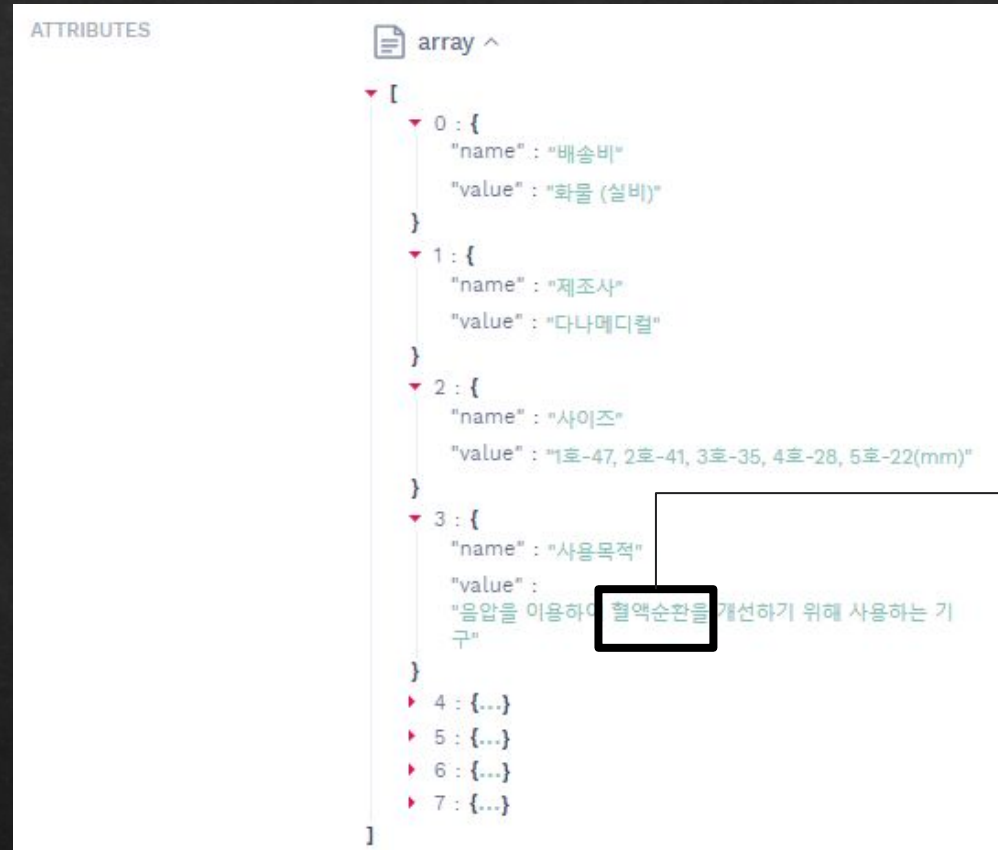
입력 키워드 : 혈액검사 -> 결과 키워드 : 부항컵...?



세팅환경 설정 (검색 필드)



검색 필드 설정



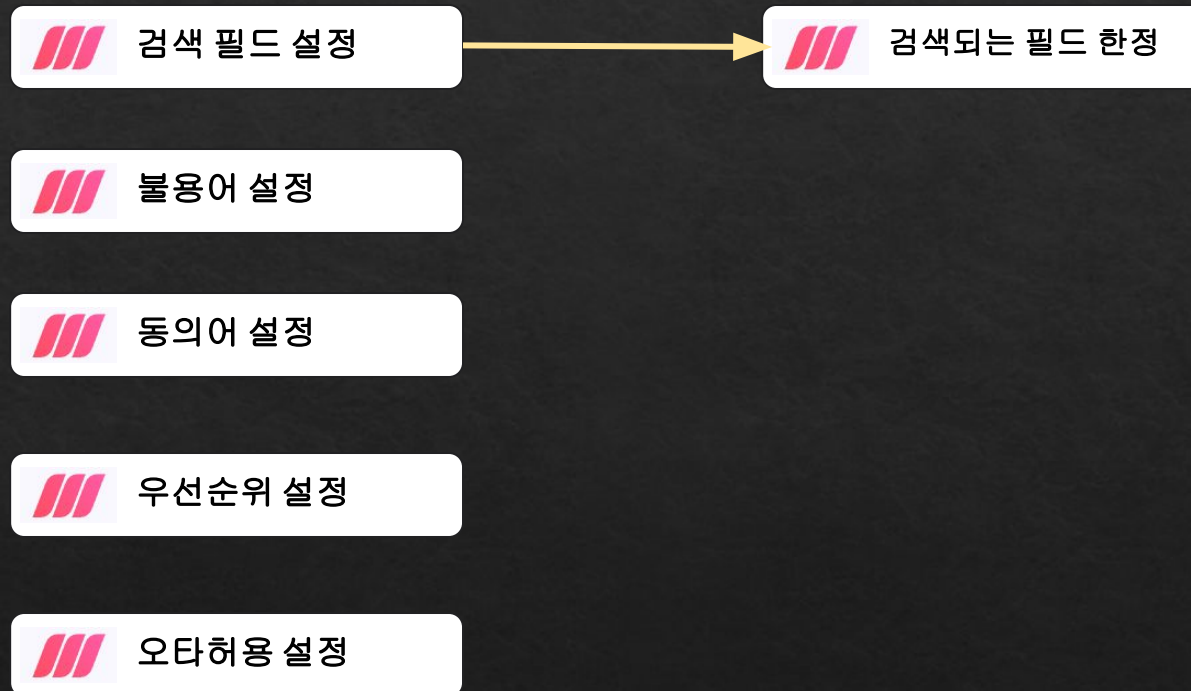
혈액순환 중 "혈액" 때문에 검색됨

모든 필드에서 검색을 하면 불필요한 정보가 너무 많다

➡ 검색 필드를 name과 description으로 한정

세팅환경 설정

세팅환경 설정



세팅환경 설정
(불용어)



불용어 설정

더모아



불용어 제거 O

0 hits

불용어 제거 X

59 hits

원인 : 불용어 중에 “더” 가 있었고, 메일리서치가 “더”는 모두 인식을 안함



불용어 미설정

세팅환경 설정

세팅환경 설정



세팅환경 설정
(동의어)



동의어 설정

Meilisearch_orgin과
동일한 환경에서 테스트하기 위해 미설정

세팅환경 설정

세팅환경 설정



세팅환경 설정 (우선순위)



우선순위 설정

Meilisearch(ogrin)



words



typo



proximity



attribute



sort



exanctness

Meilisearch(ours)



words



typo



proximity



attribute



sort



exanctness

동일한 환경에서 테스트하기 위해 미설정

세팅환경 설정

세팅환경 설정



세팅환경 설정 (오타허용)



오타허용 설정

(우리가 생각하는) 오타허용 :

메뉴얼 -> 매뉴얼

(메일리서치) 오타허용 :

메뉴얼 -> 메뉴판

동일한 환경에서 테스트하기 위해 미설정

세팅환경 설정

세팅환경 설정



작업 과정

사용자 사전 생성

빌드

세팅환경

POST 데이터

POST 데이터



Presentation INDEX |

01 결과물 소개

02 작업 과정

03 성능 평가

04 추후 진행해야 할 사항

성능 평가

Precision

Recall

F1 Score

Precision at K

성능 평가 지표

Precision

$$\frac{\text{검색된 적합 문서 수}}{\text{검색된 문서 수}}$$

Recall

$$\frac{\text{검색된 적합 문서 수}}{\text{적합 문서 수}}$$

F1 Score

$$\frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

Precision at K

검색된 문서 중 상위 K개에 포함된 적합 문서 수

성능 평가 환경

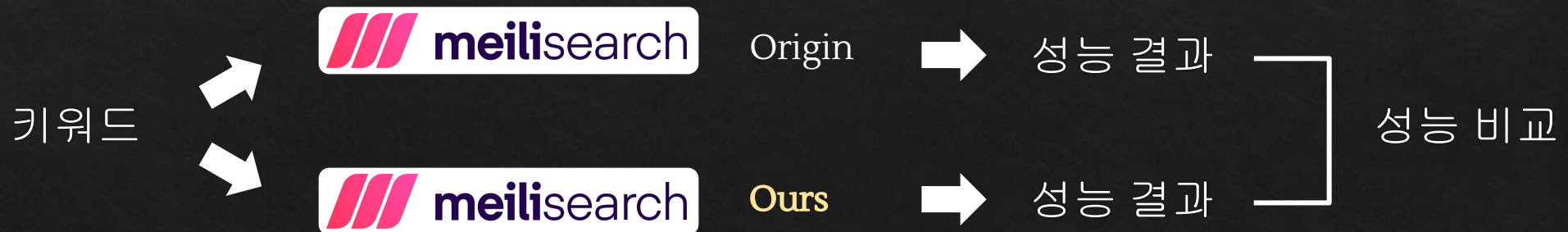
적합 문서 정의

적합 문서 정의에 따라서 성능 평가 결과가 달라짐



정의 1) NAME에 검색 키워드가 있는 문서

정의 2) NAME + DESCRIPTION에 검색 키워드가 있는 경우



성능 평가 방식



성능 평가 결과

		NAME				
			Precision	Recall	F1 Score	Precision at 10
	Origin		0.3932	0.9162	0.5503	0.2782
	Ours		0.5373	0.9751	0.6929	0.2933

성능 평가 결과

NAME + DESCRIPTION		Precision	Recall	F1 Score	Precision at 10
	Origin	0.7632	0.8851	0.7721	0.8472
	Ours	0.8726	0.9657	0.8864	0.8502

오타 허용

- 오타 허용 범위에 해당하는 다섯 글자 키워드를 검색했을 때의 예시



F1 Score



- 글자수 5 이상 (오타 허용) ➡ 0.7097
- 글자수 5 이하 (오타 허용 X) ➡ 0.9320

Presentation INDEX |

01 결과물 소개

02 작업 과정

03 성능 평가

04 추후 진행해야 할 사항

추후 진행할 사항

- 사용자 사전 세밀화
 - 대보탕 입력시, 녹용대보탕에서 "녹용+대보탕"으로 인식못하면 기존 메일리서치와 같음
- 오타허용에 대해서 추가적인 실험이 필요
- Meilisearch Korean 추가
 - 아직까지 빌드하는 과정에서 에러사항이 발생
 - 추후 활용해보는 것을 추천

Thank you



정새롬



성재훈



이태훈