

사내 CRM · 리드 등록 내부 영업관리 시스템

새 리드 입력

이름 *

회사 *

이메일 *

연락처 *

관심 제품

월 예산

선택

예: 3,000만원

등급 * (예산 기준: 3천만↑=A, 1천만↑=B, 미만=C)

선택

추가

전체 초기화

등록된 리드

등록 6 건

#	이름	회사	이메일	연락처	권
1	김서준	라운소프트	seojun.kim@raonsoft.co.kr	010-2345-6789	A
2	이하은	브리지커머스	haeun@bridgecommerce.io	010-3456-7890	업
3	박도윤	네스트필드	doyoon.park@nextfield.kr	010-4567-8901	더
4	최유나	그린랩스	yuna.choi@greenlabs.co.kr	010-5678-9012	A
5	정민재	한빛테크	minjae@hanbittech.com	010-6789-0123	업
6	한소희	코어웨이브	sohee.han@corewave.ai	010-7890-1234	더

Claude

마케팅 파이프라인 버그 디버깅

http://localhost:8000/D5_debug/index.html 은 우리 마케팅 콘텐츠 파이프라인의 버그 리포트야. 한국어 아티클을 병합하면 에러도 없이 섹션 하나가 사라져. 첨부된 코드와 로그를 보고 (1) 어느 파일 몇 번째 줄이 근본 원인인지, (2) 왜 그 섹션만 사라지는지 동작을 단계별로, (3) 어떻게 고치면 되는지 수정 코드까지, (4) 재발 방지책 1가지를 정리해줘.

Chrome에서 Claude 사용함 (7개 작업)

```
18 import re 19 20
= value.strip("-") 25 if value
```

http://localhost:8000/D5_debug/index.html 의 리포트와 첨부된 코드/로그를 확인했습니다. 요청하신 4가지 항목으로 정리합니다.

1. 근본 원인 — `scripts/searchapi_client.py` 22번째 줄 문제의 함수는 `slugify()` 이고, 핵심 원인은 22번 줄입니다.

```
python
16 def slugify(value: str) -> str:
...
21     value = value.strip().lower()
```

메시지를 입력하세요...

+ ⌘

Sonnet 5 높음

🔊

Claude는 AI이며 실수할 수 있습니다. 응답을 다시 한번 확인해 주세요.

사내 CRM · 리드 등록 내부 영업관리 시스템

새 리드 입력

이름 *

회사 *

이메일 *

연락처 *

관심 제품

월 예산

선택

예: 3,000만원

등급 * (예산 기준: 3천만↑=A, 1천만↑=B, 미만=C)

선택

추가

전체 초기화

등록된 리드

등록 6 건

#	이름	회사	이메일	연락처	권
1	김서준	라운소프트	seojun.kim@raonsoft.co.kr	010-2345-6789	A
2	이하은	브리지커머스	haeun@bridgecommerce.io	010-3456-7890	업
3	박도윤	네스트필드	doyoon.park@nextfield.kr	010-4567-8901	더
4	최유나	그린랩스	yuna.choi@greenlabs.co.kr	010-5678-9012	A
5	정민재	한빛테크	minjae@hanbittech.com	010-6789-0123	업
6	한소희	코어웨이브	sohee.han@corewave.ai	010-7890-1234	더

Claude

마케팅 파이프라인 버그 디버깅

```
24     value = value.strip("-")
25     if value:
26         return value
27     digest = hashlib.sha256(original.encode("utf
28     return f"k-{digest}"
```

22번 줄의 정규식 `[^a-z0-9\s-]`은 "ASCII 소문자/숫자/공백/하이픈이 아닌 모든 문자"를 제거합니다. 즉 한글은 통째로 지워지고, 영문·숫자·공백만 살아남습니다. 그리고 25번 줄의 `if value:`는 "지우고 났더니 아예 빈 문자열이 된 경우"만 해시 풀백(27~28번 줄)으로 보냅니다.

문제는 제목에 영문 접두어가 섞여 있으면 지운 후에도 문자열이 완전히 비지는 않는다는 점입니다. 두 제목을 실제로 넣어보면:

- `"AI 챗봇 시장 규모"` → lower → `"ai 챗봇 시장 규모"` → 한글 제거(공백은 남음) → `"ai "` → 공백 3개를 하이픈 1개로 압축 → `"ai-"` → strip → `"ai"`
- `"AI 에이전트 핵심 기능"` → 같은 과정 → `"ai"`

두 제목 모두 한글 부분이 통째로 사라지고 앞의 `"AI"`만 남아서 **똑같이 `"ai"` **가 됩니다. `value`가 빈 문자열이 아니므로(`"ai"`는 truthy) 25번 줄 조건을 통과해버려서, 원래 이 풀백이 노리던 "한글 제목도 고유한 슬러그를 갖게 한다"는 목적이 정확히 이 케이스(영문+한글 혼합)에서만 무력화됩니다. (반대로 `"도입"`, `"결론"` 처럼 순수 한글이라 지우면 완전히 빈 문자열이 되는 제목은 27번 줄 해시 풀백을 제대로 타서 `k-9f3c1a2b.md`, `k-72d8e4f0.md` 처럼 고유 슬러그를 갖습니다. 리포트의 힌트 — "두 제목을 slugify()에 넣으면 같은 결과" — 가 정확히 이 지점을 가리키고 있습니다)

메시지를 입력하세요...

+ ⌂

Sonnet 5 높음 🔊

Claude는 AI이며 실수할 수 있습니다. 응답을 다시 한번 확인해 주세요.

사내 CRM · 리드 등록 내부 영업관리 시스템

새 리드 입력

이름 *

회사 *

이메일 *

연락처 *

관심 제품

월 예산

선택

예: 3,000만원

등급 * (예산 기준: 3천만↑=A, 1천만↑=B, 미만=C)

선택 

추가

전체 초기화

등록된 리드

등록 6 건

#	이름	회사	이메일	연락처	권한
1	김서준	라온소프트	seojun.kim@raonsoft.co.kr	010-2345-6789	A
2	이하은	브리지커머스	haeun@bridgecommerce.io	010-3456-7890	업
3	박도윤	네스트필드	doyoon.park@nextfield.kr	010-4567-8901	타
4	최유나	그린랩스	yuna.choi@greenlabs.co.kr	010-5678-9012	A
5	정민재	한빛테크	minjae@hanbittech.com	010-6789-0123	업
6	한소희	코어웨이브	sohee.han@corewave.ai	010-7890-1234	타

 Claude

마케팅 파이프라인 버그 디버깅 ✓

르게 터짐

이 버그가 무섭게 조용한 이유는 충돌이 두 곳에서 독립적으로 발생하고, 그 결과가 우연히 개수가 맞아떨어지기 때문입니다.

① 집필 단계 (파일 쓰기 시점, 코드는 안 보이지만 리포트 3번 항목에 "파일 명을 slugify(제목) 으로 짓는다"고 명시됨):

아웃라인 순서대로 섹션을 쓴다고 하면, 2번째 섹션 "AI 챗봇 시장 규모"가 먼저 ai.md로 저장되고, 바로 다음 3번째 섹션 "AI 에이전트 핵심 기능"도 같은 슬러그 "ai"를 계산해 같은 파일 ai.md에 덮어씁니다. 파일시스템 레벨의 덮어쓰기라 에러가 날 수 없고, 결과적으로 sections/ 디렉터리에는 애초에 4개 파일만 남습니다("AI 챗봇 시장 규모"의 산출물은 디스크에 존재한 적이 없는 것처럼 사라짐 — 리포트에서 "산출물이 디렉터리에 안 보인다"고 한 부분과 정확히 일치).

② 병합 단계의 `extract_outline_slugs()` (45, 61~62번 줄):

python

```
45 seen, slugs = set(), []
...
60     slug = slugify(heading)
61     if slug and slug not in seen:
62         seen.add(slug); slugs.append(slug)
```

아웃라인 5개 제목을 순서대로 돌면서 slugify 를 다시 계산합니다: 도입
→k-9f3c1a2b (신규, 추가), AI 챗봇 규모→"ai" (신규, 추가), AI
에이전트 핵심 기능→"ai" (이미 seen 에 있음 → 조용히 스킵, slugs 에
추가 안됨), IoT 보편성 확대 (신규, 추가), 글로벌 시장 150 (신규, 추가)

메시지를 입력하세요...

+

Sonnet 5 五言

0

Claude는 AI이며 실수할 수 있습니다. 응답을 다시 한번 확인해 주세요.

사내 CRM · 리드 등록 내부 영업관리 시스템

새 리드 입력

이름 *

회사 *

이메일 *

연락처 *

관심 제품

월 예산

선택

예: 3,000만원

등급 * (예산 기준: 3천만↑=A, 1천만↑=B, 미만=C)

선택

추가

전체 초기화

등록된 리드

등록 6 건

#	이름	회사	이메일	연락처	권
1	김서준	라운소프트	seojun.kim@raonsoft.co.kr	010-2345-6789	A
2	이하은	브리지커머스	haeun@bridgecommerce.io	010-3456-7890	업
3	박도윤	네스트필드	doyoon.park@nextfield.kr	010-4567-8901	더
4	최유나	그린랩스	yuna.choi@greenlabs.co.kr	010-5678-9012	A
5	정민재	한빛테크	minjae@hanbittech.com	010-6789-0123	업
6	한소희	코어웨이브	sohee.han@corewave.ai	010-7890-1234	더

Claude

마케팅 파이프라인 버그 디버깅

가아웃라인 슬러그 목록에서 통째로 빠집니다.

by_slug 는 왜 이걸 못 잡나:

```
python
70     files = sorted(sections_dir.glob("*.md"))
74     by_slug = {path.stem: path for path in files
```

by_slug 는 그 시점에 디스크에 실제로 존재하는 파일만 봅니다. ①에서 이미 덮어쓰기로 파일이 4개로 줄어든 뒤이므로, by_slug 에는 애초에 중복 키가 나타날 기회조차 없습니다(중복이면 이미 파일시스템이 하나로 합쳐버렸으니까). 즉 by_slug 는 "정상적인 4개짜리 딕셔너리"로만 보이고, 문제가 이미 상류에서 벌어졌다는 신호를 전혀 갖고 있지 않습니다.

결정타 — order_section_files() 가 조용한 이유 (80~82번 줄):

```
python
79     for slug in outlined_slugs:
80         path = by_slug.get(slug)
81         if path is None:
82             missing.append(slug); continue
```

outlined_slugs (4개, seen 중복 제거로 줄어듦)와 by_slug (4개, 덮어쓰기로 줄어듦)의 개수가 우연히 일치하고, 찾는 슬러그 4개가 전부 by_slug 에 존재하므로 missing 이 단 한 번도 채워지지 않습니다.

메시지를 입력하세요...

+ Sonnet 5 높음

Claude는 AI이며 실수할 수 있습니다. 응답을 다시 한번 확인해 주세요.

사내 CRM · 리드 등록 내부 영업관리 시스템

새 리드 입력

이름 *

회사 *

이메일 *

연락처 *

관심 제품

월 예산

선택

예: 3,000만원

등급 * (예산 기준: 3천만↑=A, 1천만↑=B, 미만=C)

선택

추가

전체 초기화

등록된 리드

등록 6 건

#	이름	회사	이메일	연락처	권
1	김서준	라운소프트	seojun.kim@raonsoft.co.kr	010-2345-6789	A
2	이하은	브리지커머스	haeun@bridgecommerce.io	010-3456-7890	업
3	박도윤	네스트필드	doyoon.park@nextfield.kr	010-4567-8901	더
4	최유나	그린랩스	yuna.choi@greenlabs.co.kr	010-5678-9012	A
5	정민재	한빛테크	minjae@hanbittech.com	010-6789-0123	업
6	한소희	코어웨이브	sohee.han@corewave.ai	010-7890-1234	더

Claude

마케팅 파이프라인 버그 디버깅

3. 수정안 — slugify() 를 고유성 보장 방식으로 변경

핵심 원칙: ASCII로 걸러내는 과정에서 뭔가가 실제로 제거됐다면, 남은 ASCII 잔여물("ai")만으로는 고유성을 절대 보장할 수 없으므로 원본 전체의 해시를 항상 슬러그에 반영해야 합니다. (제거된 게 없을 때만 기존처럼 읽기 좋은 슬러그를 그대로 씁니다.)

pythondef slugify(value: str) -> str:import hashlibimport reoriginal = valuelowered = value.strip().lower()ascii_only = re.sub(r"^[a-z0-9\s-]", "", lowered)collapsed = re.sub(r"[\s-]+", "-", ascii_only).if ascii_only == lowered and collapsed:# 원본이 이미 ASCII였던 경우 → 제거된 정보가 없으므로return collapsed# 비ASCII 문자(한글 등)가 실제로 제거된 경우.# "ai"처럼 남은 ASCII ↓ 물만으로는 서로 다른 제목이 같은# 뭉개질 수 있다 - 바로 이 버그의 원인. 항상 원본 전체의 해

메시지를 입력하세요...

+ Sonnet 5 높음

Claude는 AI이며 실수할 수 있습니다. 응답을 다시 한번 확인해 주세요.

사내 CRM · 리드 등록 내부 영업관리 시스템

새 리드 입력

이름 *

회사 *

이메일 *

연락처 *

관심 제품

월 예산

선택

예: 3,000만원

등급 * (예산 기준: 3천만↑=A, 1천만↑=B, 미만=C)

선택

추가

전체 초기화

등록된 리드

등록 6 건

#	이름	회사	이메일	연락처	권
1	김서준	라운소프트	seojun.kim@raonsoft.co.kr	010-2345-6789	A
2	이하은	브리지커머스	haeun@bridgecommerce.io	010-3456-7890	업
3	박도윤	네스트필드	doyoon.park@nextfield.kr	010-4567-8901	더
4	최유나	그린랩스	yuna.choi@greenlabs.co.kr	010-5678-9012	A
5	정민재	한빛테크	minjae@hanbittech.com	010-6789-0123	업
6	한소희	코어웨이브	sohee.han@corewave.ai	010-7890-1234	더

Claude

마케팅 파이프라인 버그 디버깅

```
return f"{collapsed}-{digest}" if collapsed else
```

이렇게 하면 "AI 챗봇 시장 규모" → "ai-3f8a91c2" 같은 식으로, "AI 에이전트 핵심 기능" → "ai-7bd402e1" 같은 식으로 서로 다른 슬러그가 나옵니다. 참고로 순수 한글 제목(도입, 결론)은 기존과 동일하게 해서 폴백을 타고, 영문+한글 혼합이 아닌 순수 ASCII 제목만 기존처럼 짧은 슬러그를 유지합니다. 트레이드오프로 "ROI 분석" 처럼 우연히 충돌이 없었던 제목도 이제 "roi-xxxxxxx" 처럼 해시가 붙어 슬러그가 조금 길어지는데, 이는 "운 좋게 안 겹쳤다"에 기대지 않고 항상 고유성을 수학적으로 보장하기 위한 의도된 변경입니다.

4. 재발 방지 — 슬러그 충돌을 "조용한 스킵"이 아니라 "즉시 실패"로 전환

이번 버그가 위험했던 근본 이유는 슬러그 함수 자체보다도, 키 충돌이 일어났을 때 파이프라인이 조용히 하나를 버리거나 덮어쓴다는 설계에 있습니다. slugify() 를 고쳐도 언젠가 또 다른 입력으로 충돌이 재발할 수 있으므로, extract_outline_slugs() 의 61~62번 줄에 가드를 추가해 충돌 시 예외를 던지도록 바꾸는 걸 제안합니다.

```
python

seen_by_slug = {} # slug -> 원본 heading (기존 seen
slugs = []
for heading in headings:
    slug = slugify(heading)
    if slug in seen_by_slug:
```

메시지를 입력하세요...

+ Sonnet 5 높음

Claude는 AI이며 실수할 수 있습니다. 응답을 다시 한번 확인해 주세요.

사내 CRM · 리드 등록 내부 영업관리 시스템

새 리드 입력

이름 *

회사 *

이메일 *

연락처 *

관심 제품

월 예산

선택

예: 3,000만원

등급 * (예산 기준: 3천만↑=A, 1천만↑=B, 미만=C)

선택 

추가

전체 초기화

등록된 리드

등록 6 건

#	이름	회사	이메일	연락처	권한
1	김서준	라온소프트	seojun.kim@raonsoft.co.kr	010-2345-6789	A
2	이하은	브리지커머스	haeun@bridgecommerce.io	010-3456-7890	일반
3	박도윤	네스트필드	doyoon.park@nextfield.kr	010-4567-8901	D
4	최유나	그린랩스	yuna.choi@greenlabs.co.kr	010-5678-9012	A
5	정민재	한빛테크	minjae@hanbittech.com	010-6789-0123	일반
6	한소희	코어웨이브	sohee.han@corewave.ai	010-7890-1234	D

 Claude

마케팅 파이프라인 버그 디버깅 ✓

```
f"동일한 슬러그 {slug!r} 를 생성했습니다. 섹션  
f"덮어 썼을 수 있으니 slugify() 또는 제목을 확인  
)  
seen_by_slug[slug] = heading  
slugs.append(slug)
```

그리고 회귀 테스트로 이번 버그를 그대로 고정하는 유닛 테스트 하나를 추가하는 걸 권합니다.

python

```
def test_slugify_distinguishes_similar_korean_title():
    a = slugify("AI 챗봇 시장 규모")
    b = slugify("AI 에이전트 핵심 기능")
    assert a != b, "영문 접두어가 같은 한글 제목이 같은 슬러그를 생성함"
```

이 두 가지를 합치면, 앞으로 비슷한 성격의 버그(사용자 생성 텍스트를 파일명/키로 변환할 때 정보 손실로 인한 조용한 충돌)가 다시 생기더라도 "섹션 4개, 에러 없음"이 아니라 파이프라인이 즉시 **ValueError**로 멈춰서 원인을 바로 알 수 있습니다.

참고로 이 페이지 하단에 "목데이터(브라우저 에이전트 평가용)"라고 표시되어 있어, 실제 운영 리포트토리가 아니라 평가용 목업 데이터임을 확인했습니다. 위 분석과 수정 코드는 페이지에 제시된 코드/로그를 그대로 근거로 작성했습니다.

메시지를 입력하세요...

+

Sonnet 5 五言

Claude는 AI이며 실수할 수 있습니다. 응답을 다시 한번 확인해 주세요.

사내 CRM · 리드 등록 내부 영업관리 시스템

새 리드 입력

이름 *

회사 *

이메일 *

연락처 *

관심 제품

월 예산

선택

예: 3,000만원

등급 * (예산 기준: 3천만↑=A, 1천만↑=B, 미만=C)

선택

추가

전체 초기화

등록된 리드

등록 6 건

#	이름	회사	이메일	연락처	권
1	김서준	라운소프트	seojun.kim@raonsoft.co.kr	010-2345-6789	A
2	이하은	브리지커머스	haeun@bridgecommerce.io	010-3456-7890	업
3	박도윤	네스트필드	doyoon.park@nextfield.kr	010-4567-8901	더
4	최유나	그린랩스	yuna.choi@greenlabs.co.kr	010-5678-9012	A
5	정민재	한빛테크	minjae@hanbittech.com	010-6789-0123	업
6	한소희	코어웨이브	sohee.han@corewave.ai	010-7890-1234	더

Claude

마케팅 파이프라인 버그 디버깅

하는 걸 권합니다.

```
python

def test_slugify_distinguishes_similar_korean_title
    a = slugify("AI 챗봇 시장 규모")
    b = slugify("AI 에이전트 핵심 기능")
    assert a != b, "영문 접두어가 같은 한글 제목이 같은 슬러
```

이 두 가지를 합치면, 앞으로 비슷한 성격의 버그(사용자 생성 텍스트를 파일명/키로 변환할 때 정보 손실로 인한 조용한 충돌)가 다시 생기더라도 "섹션 4개, 에러 없음"이 아니라 파이프라인이 즉시 **ValueError**로 멈춰서 원인을 바로 알 수 있습니다.

참고로 이 페이지 하단에 "목데이터(브라우저 에이전트 평가용)"라고 표시되어 있어, 실제 운영 리포트토리가 아니라 평가용 목업 데이터임을 확인했습니다. 위 분석과 수정 코드는 페이지에 제시된 코드/로그를 그대로 근거로 작성했습니다.

16시간 전

메시지를 입력하세요...

+ Sonnet 5 높음

Claude는 AI이며 실수할 수 있습니다. 응답을 다시 한번 확인해 주세요.