

캐캐케케묵은 키워드 추출 시스템 개발기

발표자 소개

최현묵 (젤리)

Dev@git 운영

제대로 할 줄 아는 거 없음

배고픔

발표한다고 했던 거 후회중



발표에 앞서 주의사항

데이터 사이언스 그게 뭐임?

전 원래 그냥
서버 개발자입니다.

젤리 얘기는 걸러들어야 제맛!

케케묵은 키워드 추출 시스템 개발기

- 왜 만들었죠?
- 어떻게 만들었죠?
- 검증방법
- 후회되는 점
- 마무리

만들게 된 사연

데이터 부트캠프를
참석하게 됨

난 다 할 수 있어!



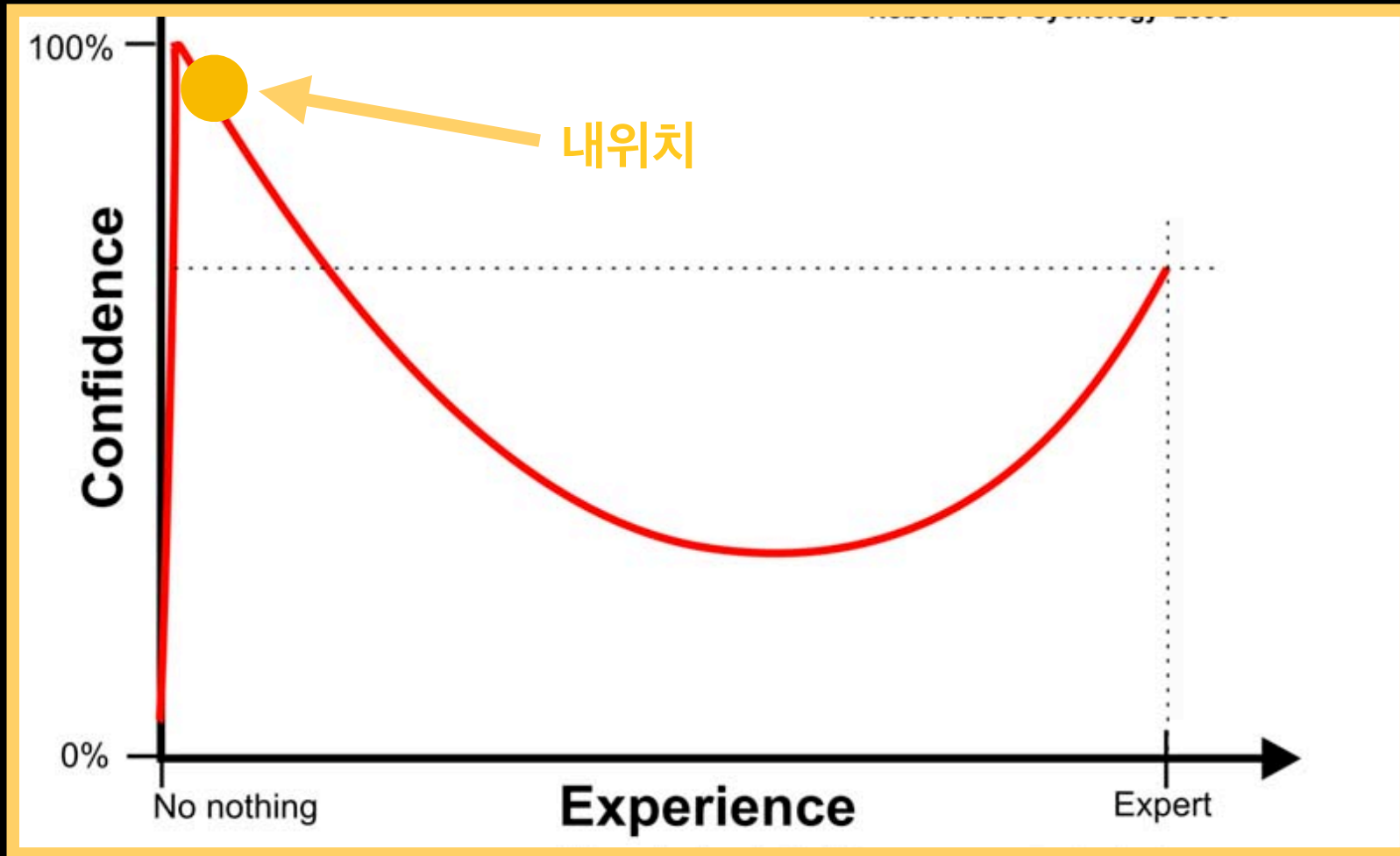
키야 데이터에 취한다

난 다 할 수 있어!



키야 데이터에 취한다

근데 통계가 뭐예요?



이쯤에서 보는 더닝 크루거

아주 그냥 정점 찍음

마침 쓰고 있던 검색엔진에서
TF 와 DF라는 걸
추출할 수 있다고 들음

그걸로 키워드 추출할 수
있을 것 같아요.

안돼!! 그런말 함부로 하는 거 아니야! 과거의 나 XX야!!!

그걸로 키워드 추출할 수
있을 것 같아요.

그래서 만들게 되었습니다.

이제부터 케케묵은 이야기를 해봅시다.

어떻게 만들어요??

TF - IDF

T(erm) F(requency)

하리보 젤리는 그냥 곰모양인 골드베렌이 제일 유명하지만, 그외에도 콜라맛 젤리, 개구리 모양 젤리, 베리모양 젤리, 스머프 모양 젤리, 열대과일맛 젤리, 복숭아맛 젤리, 달팽이 모양 젤리, 박쥐모양 젤리 등등 다양한 젤리를 판다.

“아 배고프다.. 뭐먹지.. 젤리! 뭐먹을래?”

“글썸 아무거나 맛있는 거...”

“아 그러니까 맛있는 거 뭐...”

“아 몰라. 아무튼 맛있는 거”

왼쪽이 TF가 높다

하리보 젤리는 그냥 곰모양인 골드베렌이 제일 유명하지만, 그외에도 콜라맛 젤리, 개구리 모양 젤리, 베리모양 젤리, 스머프 모양 젤리, 열대과일맛 젤리, 복숭아맛 젤리, 달팽이 모양 젤리, 박쥐모양 젤리 등등 다양한 젤리를 판다.

“아 배고파.. 뭐먹지.. 젤리! 뭐먹을래?”

“글쎄 아무거나 맛있는 거...”

“아 그러니까 맛있는 거 뭐...”

“아 몰라. 아무튼 맛있는 거”

왼쪽이 TF가 높다

하리보 젤리는 그냥 곰모양인 골드베렌이 제일 유명하지만, 그외에도 콜라맛 젤리, 개구리 모양 젤리, 베리맛 젤리, 스머프 모양 젤리, 글대과일맛 젤리, 복숭아맛 젤리, 달팽이 모양 젤리, 박쥐 모양 젤리 등등 다양한 젤리를 판다.

9/33 >

“아 배고파.. 뭐먹지.. 젤리! 뭐먹을래?”

“글쎄 아무거나 맛있는 거...” 1/18

“아 그러니까 맛있는 거 뭐...”

“아 몰라. 아무튼 맛있는 거”

D(ocument) F(requency)

하리보 젤리는 그냥
곰모양인 골드베렌이
제일 유명하지만, 그
외에도 콜라맛 젤리,
개구리 모양 젤리, 베
리모양 젤리, 스머프
모양 젤리, 열대과일
맛 젤리, 복숭아맛 젤
리, 달팽이 모양 젤리,

“아 배고파.. 뭐먹지.. 젤
리! 뭐먹을래?”

“글쎄 아무거나 맛있는
거...”

“아 그러니까 맛있는 거
뭐...”

정환
바보

정환
바보

정환
바보

정환
바보

“아 배고파.. 뭐먹지.. 젤
리! 뭐먹을래?”

“글쎄 아무거나 맛있는
거...”

“아 그러니까 맛있는 거
뭐...”

하리보 젤리는 그냥 곰
모양인 골드베렌이 제일
유명하지만, 그외에도
콜라맛 젤리, 개구리 모
양 젤리, 베리모양 젤리,
스머프 모양 젤리, 열대
과일 맛 젤리, 복숭아맛
젤리, 달팽이 모양 젤리,

“아 배고파.. 뭐먹지.. 젤
리! 뭐먹을래?”

“글쎄 아무거나 맛있는
거...”

“아 그러니까 맛있는 거
뭐...”

정환
바보

대충 DF가 5/10라고 친다

하리보 젤리는 그냥
곰모양인 골드베렌이
제일 유명하지만, 그
외에도 콜라맛 젤리,
개구리 모양 젤리, 베
리모양 젤리, 스머프
모양 젤리, 열대과일
맛 젤리, 복숭아맛 젤
리, 달팽이 모양 젤리,

“아 배고파.. 뭐먹지.. 젤
리! 뭐먹을래?”

“글쎄 아무거나 맛있는
거...”

“아 그러니까 맛있는 거
뭐...”

정환
바보

정환
바보

정환
바보

정환
바보

“아 배고파.. 뭐먹지.. 젤
리! 뭐먹을래?”

“글쎄 아무거나 맛있는
거...”

“아 그러니까 맛있는 거
뭐...”

하리보 젤리는 그냥 곰
모양인 골드베렌이 제일
유명하지만, 그외에도
콜라맛 젤리, 개구리 모
양 젤리, 베리모양 젤리,
스머프 모양 젤리, 열대
과일 맛 젤리, 복숭아맛
젤리, 달팽이 모양 젤리,

“아 배고파.. 뭐먹지.. 젤
리! 뭐먹을래?”

“글쎄 아무거나 맛있는
거...”

“아 그러니까 맛있는 거
뭐...”

정환
바보

사실 좀 틀린 설명..

그럼 다시 TF-IDF

하리보 젤리는 그냥 곰모양인 골드베렌이 제일 유명하지만, 그외에도 콜라맛 젤리, 개구리 모양 젤리, 베리모양 젤리, 스머프 모양 젤리, 열대과일 맛 젤리, 복숭아맛 젤리, 달팽이 모양 젤리, 박쥐 모양 젤리 등등 다양한 젤리를 판다.

9/33

하리보 젤리는 그냥 곰모양인 골드베렌이 제일 유명하지만, 그외에도 콜라맛 젤리, 개구리 모양 젤리, 베리모양 젤리, 스머프 모양 젤리, 열대과일 맛 젤리, 복숭아맛 젤리, 달팽이 모양 젤리,

1

“아 배고파.. 뭐먹지.. 젤리! 뭐먹을래?”

“글쎄 아무거나 맛있는 거...”

“아 그러니까 맛있는 거 뭐...”

정환

정환

정환

바보

바보

바보

5/10

정환
바보

“아 배고파.. 뭐먹지.. 젤리! 뭐먹을래?”

“글쎄 아무거나 맛있는 거...”

“아 그러니까 맛있는 거 뭐...”

3

하리보 젤리는 그냥 곰모양인 골드베렌이 제일 유명하지만, 그외에도 콜라맛 젤리, 개구리 모양 젤리, 베리모양 젤리, 스머프 모양 젤리, 열대과일 맛 젤리, 복숭아맛 젤리, 달팽이 모양 젤리,

4

“아 배고파.. 뭐먹지.. 젤리! 뭐먹을래?”

“글쎄 아무거나 맛있는 거...”

“아 그러니까 맛있는 거 뭐...”

5

정환
바보

$$9/33 * 1(5/10)$$

TF * I(nverse)DF

$$9/33 * 1(5/10)$$

$$\frac{9}{33} * \frac{10}{5} =$$
$$0.54545454\dots$$

하리보 젤리는 그냥 곰모양인 골드베렌이 제일 유명하지만, 그외에도 콜라맛 젤리, 개구리 모양 젤리, 베리모양 젤리, 스머프 모양 젤리, 열대과일 맛 젤리, 복숭아맛 젤리, 달팽이 모양 젤리, 박쥐 모양 젤리 등등 다양한 젤리를 판다.

위 문서의 젤리
키워드 점수는
 $0.5454545454\dots$



어때요?

참 쉽죠?

사실 여기까지는
실제로 쉽습니다.

하지만 건담이
출동하면 어떨까?

거!

언!

담!

책 서비스를 하다가 만난 건담

유학의거인

형이 왜
거기서 나와...?





기동전사 건담 MS 대전집 2013

전격 하비 매거진 편집부 저 | AK(에이케이 커뮤니케이션즈) 펴냄

★★★★★ 9.0 회원리뷰(1건) | 판매

정가 34,000원

판매가 **30,600원** (10% 할인)

YES포인트 1,700원 (5% 적립)

5만원이상 구매 시 2천원 추

결제혜택 카드/간편결제 혜택을 확인하세요

구매 시 참고사항 현재 새 상품은 구매 할 수 없습니다

기동전사 건담

기동전사 건담 THE ORIGIN

MSV

MSV-R

기동전사 건담 MS IGLOO ?1년전쟁 비록-

기동전사 건담 MS IGLOO ?목시록 0079-

기동전사 건담 MS IGLOO 2 ?중력전선-

기동전사 건담 MS IGLOO 603

기동전사 건담 제08 MS 소대

기동전사 건담 0030 주머니 속의 전쟁

기동전사 건담 팬텀 볼릿

기동전사 건담 0033 STARDUST MEMOR M-MSV(오오카와라 쿠니오 컬렉션)

기동전사 Z건담

Z-MSV

ADVANCE OF Z 티탄즈의 깃발 아래

ADVANCE OF Z 시간에 저항하는 자

기동전사 건담 에콜 듀 시엘

기동전사 건담 ZZ

건담 센티넬

기동전사 건담 역습의 샤아

더블 페이크 언더 더 건담

CCA-MSV

기동전사 건담 0083 STARDUST MEMO

기동전사 Z건담

Z-MSV

ADVANCE OF Z 티탄즈의 깃발 아래

ADVANCE OF Z 시간에 저항하는 자

기동전사 건담 에콜 듀 시엘

기동전사 건담 ZZ

건담 센티넬

기동전사 건담 역습의 샤아

더블 페이크 언더 더 건담

CCA-MSV

M-MSV(오오카와라 쿠니오 컬렉션)

기동전사 건담 UC

기동전사 건담 UC MSV

기동전사 건담 섬광의 하사웨이

기동전사 건담 F91

F91-MSV

기동전사 건담 F90

기동전사 건담 F91 포물러 전기 0122

기동전사 건담 슬루엣 포물러 91

기동전사 V건담

기동전사 V건담 MSV

기동전사 건담 SEED

기동전사 건담 SEED MSV

기동전사 건담 SEED ASTRAY

기동전사 건담 SEED DESTINY

기동전사 건담 SEED DESTINY MS

기동전사 건담 SEED DESTINY AS

기동전사 건담 SEED C.E.73 STAF

기동전사 건담 SEED C.E.73 ΔAST

기동전사 건담 SEED FRAME AST

기동전사 건담 SEED VS ASTRAY

기동전사 건담 00

기동전사 건담 00 세컨드 시즌

기동전사 건담 00 -A wakening of th

기동전사 건담 00P

기동전사 건담 00F

기동전사 건담 00I/ 00I 2314

기동전사 건담 00V

기동전사 건담 00N

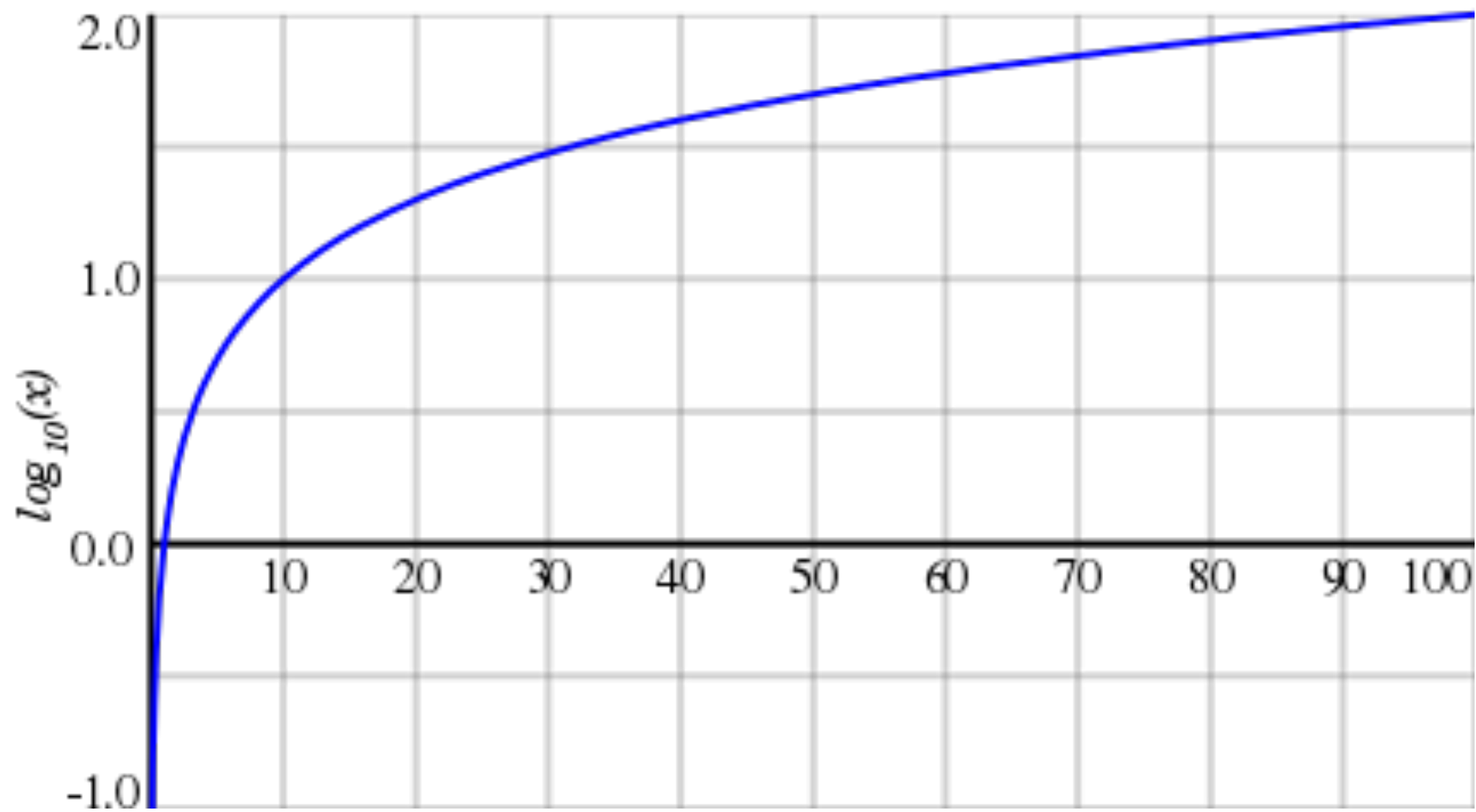
기동전사 건담 AGE

기동전사 건담 AGE EXA-LOG

기동전사 건담 AGE Unknown Sold

log

L(ord) O(h my) G(od)



다양한 TF 가중치

binary	0, 1
raw count	$f_{t,d}$
term frequency	$f_{t,d} / \sum_{t' \in d} f_{t',d}$
log normalization	$\log(1 + f_{t,d})$
double normalization 0.5	$0.5 + 0.5 \cdot \frac{f_{t,d}}{\max_{\{t' \in d\}} f_{t',d}}$
double normalization K	$K + (1 - K) \frac{f_{t,d}}{\max_{\{t' \in d\}} f_{t',d}}$

수식 복사가 안되서 부득이 하게 캡처한 점 죄송합니다.

다양한 DF 가중치

weighting scheme	IDF weight ($n_t = \{d \in D : t \in d\} $)
unary	1
inverse document frequency	$\log \frac{N}{n_t} = -\log \frac{n_t}{N}$
inverse document frequency smooth	$\log \left(1 + \frac{N}{n_t} \right)$
inverse document frequency max	$\log \left(\frac{\max_{\{t' \in d\}} n_{t'}}{1 + n_t} \right)$
probabilistic inverse document frequency	$\log \frac{N - n_t}{n_t}$

대부분의 위키 문서가 그렇지만
TF-IDF 에 관한 설명은 꼭 영문위키를 참조하세요!
한국어보다 훨씬 많은 정보를 알 수 있습니다.

추천 TF-IDF

weighting scheme	document term weight	query term weight
1	$f_{t,d} \cdot \log \frac{N}{n_t}$	$\left(0.5 + 0.5 \frac{f_{t,q}}{\max_t f_{t,q}}\right) \cdot \log \frac{N}{n_t}$
2	$1 + \log f_{t,d}$	$\log\left(1 + \frac{N}{n_t}\right)$
3	$(1 + \log f_{t,d}) \cdot \log \frac{N}{n_t}$	$(1 + \log f_{t,q}) \cdot \log \frac{N}{n_t}$

대부분의 위키 문서가 그렇지만
TF-IDF 에 관한 설명은 꼭 영문위키를 참조하세요!
한국어보다 훨씬 많은 정보를 알 수 있습니다.

사용자에게 보여줄만한
키워드만 뽑기

정말 키워드다운(?) 단어들

유저에게 보여줄만한
키워드만 뽑기

올바른 표현은 아니지만 이보다 더 좋은 표현이 없었다..

노가가다

이것도 어떻게 보면 기술력...

문서 전체에서
출현 횟수가 높은 단어부
터
내림차순으로 정렬후
선택하여 정제

형태소 분석이 난해한 한국어특성

형태소 분석에만 의지하기가 좀 껄끄러웠음

즉 사람이 직접 고른다!

몇몇 단어 같지 않은 단어들..

내림차순으로 하는 이유는
자주 사용하는 단어 먼저 처리하기 위함

키워드를 추천에
사용할 수 있을까?

각 유저의 키워드를
뽑을 수는 없을까?

카테고리 키워드도?

유저의 키워드 기준으로
카테고리 연관성을 가지고
추천해주면?

유저(카테고리)의 키워드
는 어떻게 뽑을 수 있을까?

TF-IDF 처럼 다른 유저에
게
나타나지 않는
키워드를 뽑으면?

유저들



난 자전거를 좀 좋아해



내 인생이 곧 자전거야,
난 진짜 자전거 엄청 좋아해
롱보드는
타긴하는데 별로야



난 밖에서 운동할 바에
게임을 하겠어



나도 자전거 좋아해

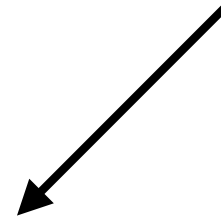


나도 자전거 좋아!



여긴 자전거 좋아하는 사람들 많네
나도 자전거 좋아

누가봐도 자전거 제일 좋아하는 애



난 자전거를 좀 좋아해



내 인생이 곧 자전거야,
난 진짜 자전거 엄청 좋아해
롱보드는
타긴하는데 별로야



난 밖에서 운동할 바에
게임을 하겠어



나도 자전거 좋아해



나도 자전거 좋아!



여긴 자전거 좋아하는 사람들 많네
나도 자전거 좋아

자전거 출현수



난 **자전거**를 좀 좋아해



내 인생이 곧 **자전거**야,
난 진짜 **자전거** 엄청 좋아해
롱보드는
타긴하는데 별로야



난 밖에서 운동할 바에
게임을 하겠어



나도 **자전거** 좋아해



나도 **자전거** 좋아!



여긴 **자전거** 좋아하는 사람들 많네
나도 **자전거** 좋아

롱보드 출현수



난 자전거를 좀 좋아해



내 인생이 곧 자전거야,
난 진짜 자전거 엄청 좋아해
롱보드는
타긴하는데 별로야



난 밖에서 운동할 바에
게임을 하겠어



나도 자전거 좋아해



나도 자전거 좋아!



여긴 자전거 좋아하는 사람들 많네
나도 자전거 좋아

TF-IDF 같은 방법으로하면
출현 수가 적은 키워드가
뵈히게 됨

즉 해당
유저만의 키워드

우리가 뽑고 싶은 건
유저의 키워드

생각해보면 우린 이미
TF-IDF로 문서의 키워드를
이미 점수로 가지고 있음

그냥 유저가 가진 문서(책)의
키워드 TF-IDF를 합산하면
유저의 키워드가 나오지 않을
까?

개발자를 위해

SQL로 표현하면?

```
SELECT SUM(score)
FROM
document_keywords
GROUP BY keyword
```

score는 keyword의 TF-IDF

이때 document_keywords 는 유저가 가진 문서라고 친다.

같은 방법으로
카테고리도 처리

하지만 건담이
출동하면 어떨까?

거!

고만해
미친놈아!!

와
폭
강
충
격

기동전사 건담

기동전사 건담 THE ORIGIN

MSV

MSV-R

기동전사 건담 MS IGLOO ?1년전쟁 비록-

기동전사 건담 MS IGLOO ?목시록 0079-

기동전사 건담 MS IGLOO 2 ?중력전선-

기동전사 건담 MS IGLOO 603

기동전사 건담 제08 MS 소대

기동전사 건담 0030 주머니 속의 전쟁

기동전사 건담 팬텀 불릿

기동전사 건담 0033 STARDUST MEMOR M-MSV(오오카와라 쿠니오 컬렉션)

기동전사 Z건담

Z-MSV

ADVANCE OF Z 티탄즈의 깃발 아래

ADVANCE OF Z 시간에 저항하는 자

기동전사 건담 에콜 듀 시엘

기동전사 건담 ZZ

건담 센티넬

기동전사 건담 역습의 샤아

더블 페이크 언더 더 건담

CCA-MSV

기동전사 건담 0083 STARDUST MEMO

기동전사 Z건담

Z-MSV

ADVANCE OF Z 티탄즈의 깃발 아래

ADVANCE OF Z 시간에 저항하는 자

기동전사 건담 에콜 듀 시엘

기동전사 건담 ZZ

건담 센티넬

기동전사 건담 역습의 샤아

더블 페이크 언더 더 건담

CCA-MSV

M-MSV(오오카와라 쿠니오 컬렉션)

기동전사 건담 JC

기동전사 건담 JC MSV

기동전사 건담 섬광의 하사웨이

기동전사 건담 F91

F91-MSV

기동전사 건담 F90

기동전사 건담 F91 포물러 전기 0122

기동전사 건담 실루엣 포물러 91

기동전사 V건담

기동전사 V건담 MSV

기동전사 건담 SEED

기동전사 건담 SEED MSV

기동전사 건담 SEED ASTRAY

기동전사 건담 SEED DESTINY

기동전사 건담 SEED DESTINY MS

기동전사 건담 SEED DESTINY AS

기동전사 건담 SEED C.E.73 STAF

기동전사 건담 SEED C.E.73 ΔAST

기동전사 건담 SEED FRAME AST

기동전사 건담 SEED VS ASTRAY

기동전사 건담 00

기동전사 건담 00 세컨드 시즌

기동전사 건담 00 -A wakening of th

기동전사 건담 00P

기동전사 건담 00F

기동전사 건담 00I/ 00I 2314

기동전사 건담 00V

기동전사 건담 00N

기동전사 건담 AGE

기동전사 건담 AGE EXA-LOG

기동전사 건담 AGE Unknown Sold

만화의 카테고리가 기동전사랑 건담?

너무 점수가 큰 **기동전사**

아웃라이어에 대해서 처리해야함

DF는 낮고 TF는 높은 경우에는
문서하나의 키워드가 카
테고리의 키워드가 되어버림

그냥 **기동전사**를
키워드에서 빼버릴까?

만약에 다른 아웃라이어가 생기면??

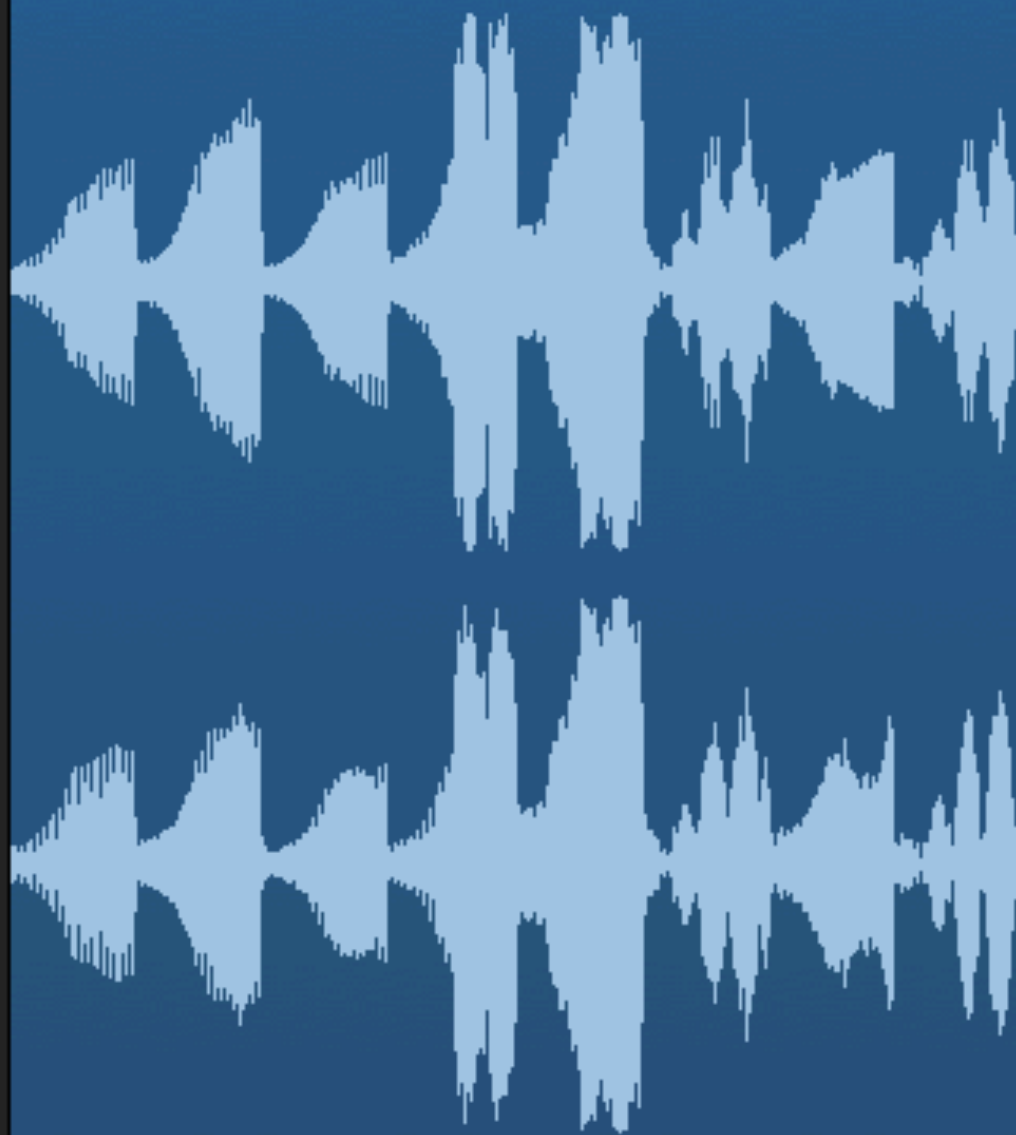
그냥 **기동전사**를
키워드에서 빼버릴까?

보편적인 방법이 필요

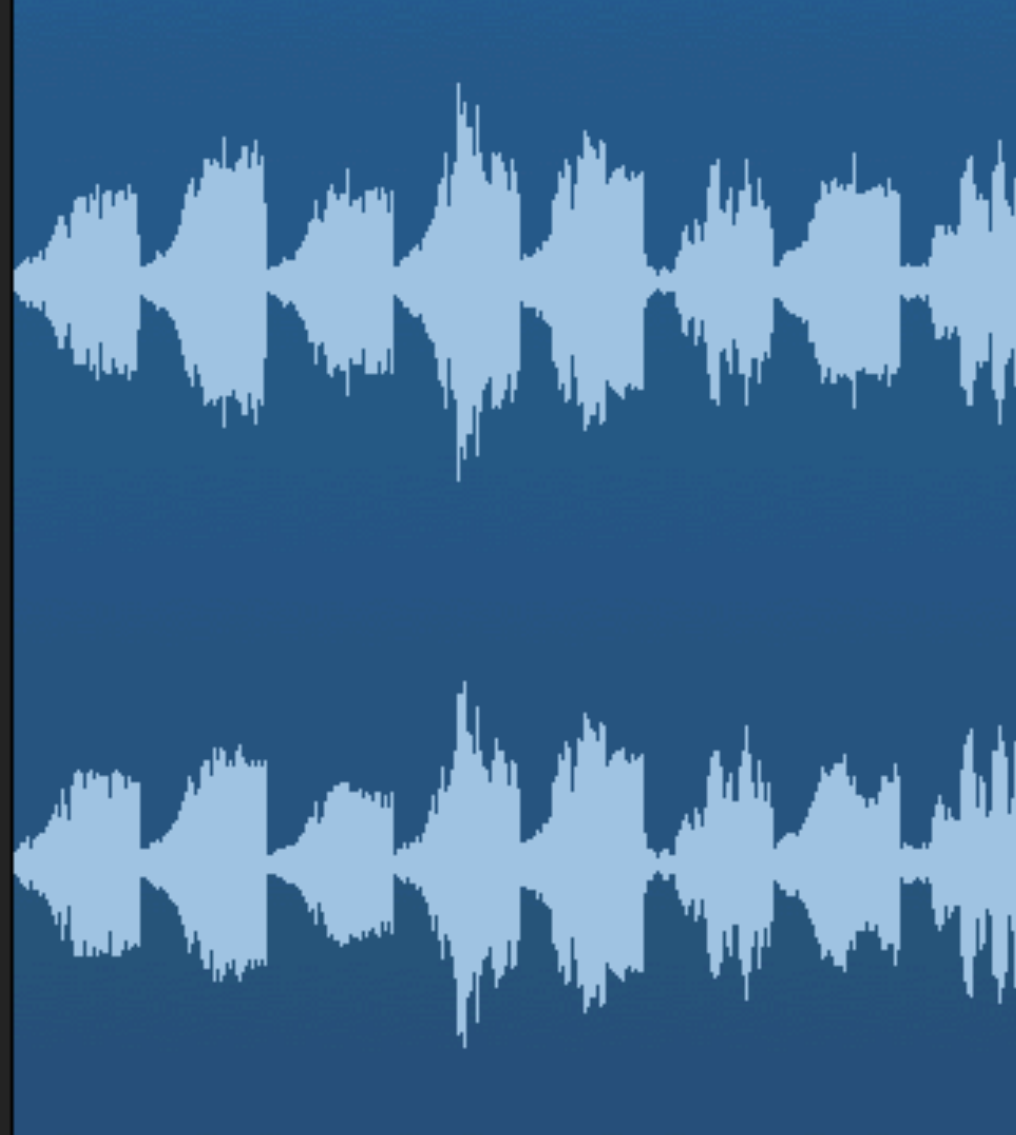
이때 생각난
오디오 컴프레서

소리의 증폭을 적게 하거나
큰 소리를 줄임으로써
오디오 신호의 **다이내믹 레**
인지를 축소시키는 장치이다.

Before compression ΩΩ



After compression ∞∞



점수를 글씨 크기로 두고
키워드를 시각화 해봅시다.



보노보노 보노보노 보노보노 보노보노

기동전사



기동전사



보노보노 보노보노 보노보노

보노보노 보노보노





보노보노

보노보노

보노보노

보노

보노



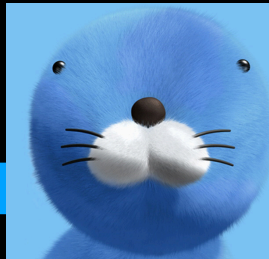
기동전사
기동전사



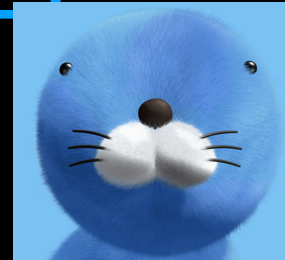
보노보노

보노보노

보노보노



보노보노



그런데 어떻게?

그런데 어떻게?

이렇게 눌러주는 거
전에 한번 봤잖아?

또 한번

log

L(ord) O(h my) G(od)

대부분 점수가 1점 이하이기 때문에
음수가 되어버림

제가 사용한 공식
 $\text{sum}(\log(K + \text{score}))$

$K(0.1 \sim 0.9)$ 값을 이용해 적절히 조정

하지만 카테고리에 속한
문서 수에 따라 점수가
천차만별로 차이날 수 있
음

사실 키워드 뽑는 건 상관 없지만,
나중에 추천엔진을 돌리기 위해 필요

점수의 범위를 설정

100점으로 맞추고 싶다면,
합산한 **최대값**을 찾은 뒤
모든 점수에 **최대값**으로
나누고 그 값에 **100**을 곱함

이렇게 하면 0~100 에 잘 분포된다!

$$\text{score} / \text{max_score} * 100$$

이제 유저가 가진 키워드와
카테고리 키워드의 점수를 조합
해
추천을 돌려줍니다!

"아 근데 저기
진짜 키워드 잘 뽑힌 건지
어떻게 확인해요?"

솔직히 전 확인을 못했어요...

True / False
Positive / Negative

점쟁이가 말한게 발생하면?

True / Positive

점쟁이가 말한게 안발생하면?

False / Positive

점쟁이가 안말한게 발생하면?

False / Negative

점쟁이가 안말한게 안발생하면?

True / Negative

점쟁이가 말한게 발생하면?

True / Positive

점쟁이가 말한게 안발생하면?

False / Positive

점쟁이가 안말한게 발생하면?

False / Negative

점쟁이가 안말한게 안발생하면?

True / Negative

점쟁이가 말한게 (XNOR) 발생하면?

True / Positive

점쟁이가 말한게 (XNOR) 안발생하면?

False / Positive

점쟁이가 안말한게 (XNOR) 발생하면?

False / Negative

점쟁이가 안말한게 (XNOR) 안발생하
면

True / Negative

검색했을 때 내가
찾고 싶은 게 검색이 안 된다면?

검색했을 때 내가
찾고 싶은 게 검색이 안 된다면?

False / Positive

검색했을 때 내가
찾고 싶은 게 검색이 안 된다면?

(NXOR)

False / Positive

검색했을 때 내가
찾고 싶은 게 검색이 안 된다면?

False / Positive

그럼 이걸 키워드에 적
용

내게 맞는 단어가 키워드 리스트에 있으면?

True / Positive

내게 안 맞는 게 키워드 리스트에 있으면?

False / Negative

두개 가지고 **정밀도**를 구할
수 있다고 생각하실 수 있
지만...

재현율

= (True / Positive)

(True / Positive) + (False / Negative)

정밀도

= (True / Positive)

(True / Positive) + (False / Positive)

정밀도

= (True / Positive)

(True / Positive) + (False / Positive)

False / Positive에 해당하는
내게 적합하지만
키워드 리스트에는 없는
키워드를 알 방법이 없음..

판단은 개인의 몫

북 스타트업 대표의 키워드

경제

세계

데이터

분석

비즈니스

철학

사람

빅데이터

마케팅

사회

인문학

과학

기업

자본

경영

인간

경제학

시장

자유

디지털

강의

소셜

구글

사랑

정치

지배

정신

역사

진화

예측

문화

결정

기술

고객

독서

도덕

자본주의

시대

시간

젤리의 키워드

자바

데이터베이스

독서

스크립트

프로그래밍

클래스

함수

애플리케이션

객체

패턴

html

서버

모듈

개발자

소프트웨어

버전

호출

설치

파이썬

css

api

문자열

생성

소크라테스

메서드

성공

매개변수

구현

virtual

카네기

프로젝트

사람

시간

클라이언트

고전

연산자

후회되는 점

사실 유저기반으로 할만한
데이터가 적어서 어쩔수 없었지만

애초에 추천은
텍스트 기반으로 하지 말자

Content-based는 어렵습니다.

- 사전에 현업지식도 꽤 있어야 하고
- 개별 콘텐츠의 메타데이터도 거의 없고
- 상품에서 그것을 뽑아내기도 어렵습니다.

이걸 나중에 봤어요..

하용호님 Deview 2015 발표중

사실 유저데이터가 없으면 CF가 힘들어서
데이터 기반으로 하기도 했어요.

키워드를 잘 뽑아낸다 한들,
대부분 Collaboration
Filtering 이 잘 동작합니다.

예를 들어, 기술 서적을 키워드만 가지고 파악한다면,
트렌드가 지난 오래된 책들이 나올 수가 있겠죠

훨씬 더 현대적인 API

Solr 대신 Elasticsearch 썼으면..

반 강제적 자바...

질문

마무리

감사합니다