

# 검색서비스 동향과 실시간 웹

2010년 2월

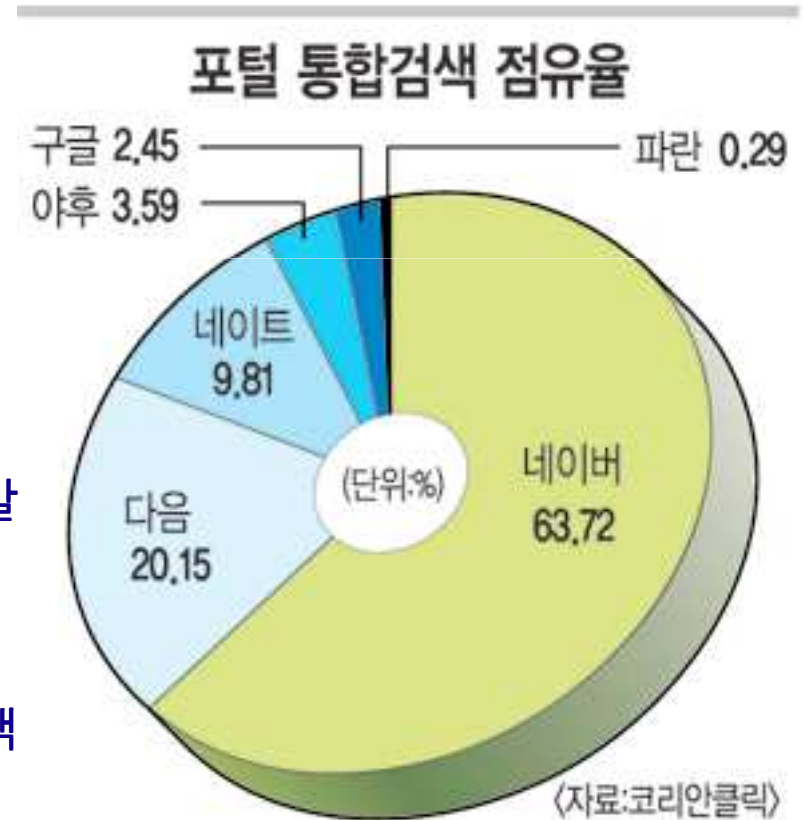
박종헌  
서울대학교 산업공학과  
[jonghun@snu.ac.kr](mailto:jonghun@snu.ac.kr)

# 해외 검색 서비스 동향



# 국내 검색 서비스 동향

- 2001 **NAVER** 통합검색
- 2002 **NAVER** 지식검색
- 2003 **empas** 지식검색
- 2004 **Google** 한국어 검색
- 2005 **dallm** 신지식 검색 **empas** 열린검색
- 2006 블로그, 카페, 게시판 검색
- 2007 동영상검색 **dallm** 자체검색엔진 개발
- 2008 **empas NATE** 합병
- 2009 **NATE** 시맨틱 검색 **dallm** 소셜 검색
- 2010 **LIVEK** 실시간 검색  
살아있는 검색



# 국내 검색 시장의 특징

포털 = 검색

짧은꼬리 형태의 검색어 분포

상태 질의어 비중: **50+%**

뉴스 검색의 비중이 높음

실시간 검색 시장 태동

DYNAMIC  
KOREA

순위	검색어	변동폭
1	안도미키	↑ 25
2	yf소나타	↑ 41
3	고양이성형	NEW!
4	이진욱	NEW!
5	메탈레이지	NEW!
6	조슬린와일튼스...	NEW!
7	나경은	↑ 1457
8	수애	↑ 13
9	유재석	NEW!
10	최지우	NEW!
11	현대자동차	NEW!
12	진재영	↑ 6199
13	나경원	↑ 273
14	대성모의고사	↑ 9230
15	cd브레이커	↓ 13
16	강혜정	↑ 79
17	강인	↓ 16
18	주아민	↑ 159
19	재범	↓ 8
20	해피투게더	↑ 1447
21	셀트리온	↑ 14
22	jyp	↑ 136
23	한화채움	↑ 31
24	원피스557화	↑ 7
25	이승기	NEW!
26	구은애	↓ 22
27	김영민	↑ 4827
28	고양이학대	↓ 25
29	에프엑스	↓ 9
30	육지혜	↑ 244



# 실시간 웹

사람들이 만든 콘텐츠의 스트림

실시간웹 >> 트위터

“지금” 무엇이 일어나고 있는가?

커뮤니케이션과 어텐션의싸이클

언제 어디서나 정보를 생성 및  
소비



# 실시간 검색



시간 가치가 있는 정보를 적시에 “발견”

스포츠 경기 결과  
연예인 소식  
사건, 사고  
신제품  
가격 정보

홈페이지  
조리법  
옛날 신문기사  
법령, 법규  
프로필

정보의 시간 의존성



# 실시간 검색 시장

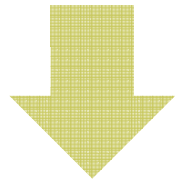


**\$70 to \$100B market**



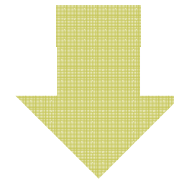
# 검색 경쟁 요소의 패러다임 변화

과거를 검색  
“Results 1 - 10 of about  
1,180,000,000 for ...”



현재를 검색  
“5 seconds ago for ...”

찾는 검색 (pull)



찾아주는 검색 (push)



# 실시간 검색의 예



# 실시간 웹 생태계

시간성  
컨텐츠



정보형  
컨텐츠

소셜  
커뮤니케이션

검색  
추천  
필터링  
병합

...

스트림

컨텐츠  
생성



twitter

Home Profile Find People Settings Help Sign out

실시간웹 컨퍼런스. 첫강의에 트윗, 페이스북 ☆  
얘기하고나면 다음사람들은 무슨얘기할  
까요? 궁금해서 참가신청. 괜히 되도않는  
얘기 할거면 차라리 노래를 불러달라는..

about 13 hours ago from TwitBird iPhone

↩ Reply ↻ Retweet



# 실시간 검색의 주요 기술적 이슈

108 Comments

f Share 62

207  
tweets

retweet

Name Yuna Kim

7 82,537  
following followers

1,359  
subscribers

-25%  
engagement

Following



친구

1,343명의 친구 전체 보기



컨텐츠 품질 측정

중복 및 스팸 처리

정보성 컨텐츠 파악

정보 신선도

컨텐츠 랭킹



# 실시간 검색에서의 콘텐츠 스코어링

$$\sigma(q, d, t) = f(\text{rel}(q, d), \text{qual}(d), \text{soc}(q, d, u), \text{age}(d, t))$$



# 실시간 검색 기반기술

## 플랫폼

- XMPP
- PubSubHubBub
- RSS cloud
- tornado, comet, ...

## 정보 수집

- 스트림 발견
- 정보 수집 자원 사용 최적화
- 정보 신선도 극대화
- 중복, 스팸 처리



## 정보 저장

- in-memory indexing
- 메모리 구조에 최적화
- 대용량 분산 처리

## 정보 처리

- 필터링
- 품질 측정
- 스트림프로파일링
- 스트림 분류
- 스코어링



# 실시간 웹에서의 기회

Time-Sensitive Marketing

Gossip Monitoring

Brand Monitoring

Alerts

Social App

On-the-Fly Tabloid

Categorical Streams

Real-Time Media Streams

First Event Detection

Hot Topic Tracking

Predictions

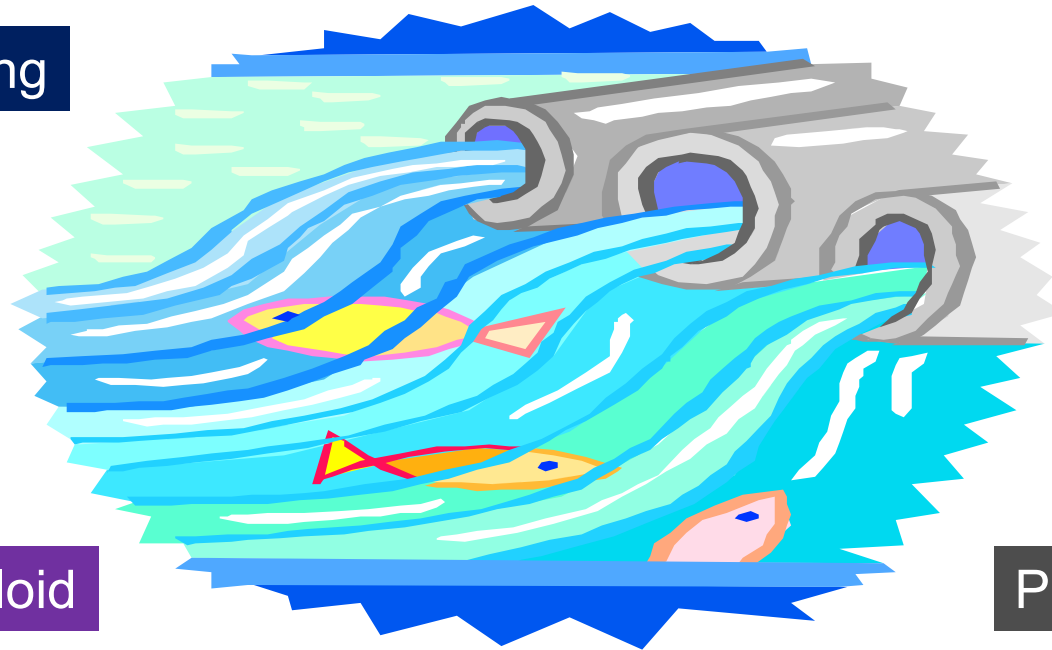
Trends

Mobile

Personalization

Recommendations

Location Awareness



# 2010년 국내 검색 시장의 키워드



◆애플 아이폰

◆구글 G1

◆삼성전자 T음니아



- 정보 허브로서의 포털 의존도 감소
- 기존 포털 사업자의 지배력 약화
- 네이버 선호도:  
**76%(PC)->67%(스마트폰)**

- 실시간 스트림의 빠른 증가
- 모바일 시장의 큰폭 성장
- 검색 시장의 경쟁 가속화
- 성공적인 해외서비스 모형의 국내 도입 시간차