hci+d lab.

Week 12 • 데이터 저널리즘

로봇 저널리즘: 뉴스 패러다임의 전환

Joonhwan Lee
human-computer interaction + design lab.

KBO 프로야구 뉴스 기사

문학구장을 뜨겁게 달군 SK와 KIA의 불방망이
(2017-07-05, KIA 17 對 18 SK, 문학)

SK는 문학구장에서 열린 KIA와의 2017 타이어뱅크 KBO리그 경기에서 38안타 10홈런 35득점을 합작하는 타격쇼 끝에 승리하였다 SK는 4개의 홈런을 기록하면서 불방망이를 뽐냈다. 다이아몬드와 팻딘은 각각 6실점과 8실점을 기록하며 부진하였다.

1회 초 KIA는 2사 2루 상황에서 최형우의 볼넷으로 2사 1, 2루 상황을 만들었으나 이후 안치홍의 3루수 라인드라이브 아웃으로 공수교대가 이루어지며 주자를 불러들이는 데에는 실패했다.

1회 말 SK는 1사 2루 상황에서 정의윤의 1타점 적시타와 한동민의 2점 홈런으로 점수 차를 벌렸으며 김동엽의 1점 홈런으로 점수 차를 벌렸다.

2회 초 KIA는 1사 1, 2루 득점찬스를 맞이하였으나 한승택의 삼진과 이명기의 좌익수 플라이로 추격에 실패했다.

3회 말 SK는 1사 1, 2루 상황에서 한동민의 1타점 적시타와 김동엽의 유격수 땅볼로 점수 차를 벌렸으며 로맥의 2점 홈런으로 점수 차를 벌렸다. 2사 1, 2루 상황에서 노수광의 투수 땅볼로 아웃카운트를 헌납해 추가득점에 실패했다.

continue…

KBO 프로야구 뉴스 기사

5회 초 KIA는 최형우의 2점 홈런으로 점수 차를 좁혔다. 이범호의 3점 홈런으로 점수 차를 좁혔고, 신종길의 2루타로 무사 2루 상황을 만들고 최원준의 1타점 2루타로 점수 차를 좁혔고, 폭투로 무사 3루 상황을 만들고 이명기의 2점 홈런으로 점수 차를 좁혔다. 버나디나의 2점 홈런으로 점수 차를 좁혔다. 무사 2, 3루 상황에서 나지완의 유격수 땅볼로 승부를 제자리로 돌려세웠고, 상대 문광은의 폭투로 2점 앞서나가기 시작했다. 그 후 7회 초에는 2사 1, 2루 상황에서 이범호의 1타점 적시타로 점수 차를 벌렸다.

8회 말 SK는 무사 1, 2루 상황에서 이재원의 2타점 2루타로 2점을 만회했으며 폭투로 2사 3루 , 김성현의 볼넷으로 2사 1, 3루 , 노수광의 볼넷으로 2사 만루 상황을 만들고 나주환의 3타점 3루타로 승부를 뒤집었으며 상대 임창용의 폭투로 점수 차를 벌렸다.

9회 초 KIA는 나지완의 2점 홈런으로 2점을 따라잡았으나 이미 승부는 기울어진 후였다.

끝내 SK는 극적인 역전승을 보여줬다. 오늘 경기의 결과 SK는 현재 3위(승률 0.563) 기록 중이다. 한편 KIA는 7연승에서 저지당해 아까워했고 80번째 경기에서 52승 28패 0무를 기록 중이며 현재 1위(승률 0.650)이다. 팀은 21안타 6홈런 4볼넷 16타점 17득점으로 노력하는 모습을 보여줬다.

Workspace
Navigate
Commands
Collaborate
Outline
Debugger
FAVORITES
controllers - /app
concerns
api_controller.rb
application_controlle
articles_controller.rb
data_provider_contr
events_controller.rb
games_controller.rb
models - /app
views - /app
api
articles
_form.html.erb
edit.html.erb
index.html.erb
index.json.jbuilde
new.html.erb
show.html.erb
show.json.jbuilde
data_provider
events
games
_form.html.erb
edit.html.erb
index.html.erb
index.json.jbuilde
new.html.erb
show.html.erb
show.json.jbuilde
layouts
application.html.e
helpers - /app
api_helper.rb
application_helper.rt
article_p3.rb
articles_helper.rb
articles_helper_back
articles_herpler_rb_
articles_herpler_rb_
articles_herpler_rb_
data_provider_helpe
events_helper.rb
events_controller
articles_controller
application.html.e
1 class ArticlesController < ApplicationController
2 before_action :set_article, only: [:show, :edit, :update, :destroy]
3 include ArticlesHelper
4 # GET /articles
5 attr_accessor :article
6 # GET /articles.json
7 def index
8 #@articles = Article.all.sort{|a,b| a.game.game_code <=> b.game.game_code }.reverse
9 @this_season = Season.where(id: 7).take
10 @games = @this_season.games
11 .select{|g| (g.able || (g.game_date <= Time.zone.now.to_date)) && (!g.is_canceled)}
12 .sort{|b,a| a.game_code <=> b.game_code}
13 @articles = @this_season.articles.sort{|a, b| a.game.game_date <=> b.game.game_date
14 @article = Article.new
15 end
16
17 # GET /articles/1
18 # GET /articles/1.json
19 def show
20 end
21
22 # GET /articles/new
23 def new
24 @games = Season.find(5).games.where(able: true)
25 .sort{|a,b| a.game_code <=> b.game_code }.reverse
26 @article = Article.new
27 end
28
29 # GET /articles/1/edit
30 def edit
31 end
32
33 # POST /articles
34 # POST /articles.json
35 def create
36 @article = Article.new(article_params)
37
38 @game = @article.game
7:12 Ruby Spaces: 2
[B] https://baseba
https://baseball-renewal-drater05.c9users.io/
Browser
Robot Journalism for KBO, ver.β(22)
경기를 선택하십시오
Create Article
Article was successfully created.
5
서동욱의 내야안타, KIA의 승리 결정지어
3
(2017-07-06, KIA 5 對 3 SK, 문학)
20170707_02:05
6일 문학구장에서 열린 2017 타이어뱅크 KBO리그 KIA와 SK의 경기에서 KIA가 서동욱의 적시 내야안타 기록에 힘입어 5:3으로 승리했다. 3:3으로 경기 중이던 8회 초 2사 1, 3루에서 서동욱의 적시 내야안타가 1득점을 얻으며 승기를 잡았다. 오늘 승리에 결정적 기여를 한 서동욱은 시즌 206타수 63안타 5홈런 20볼넷 38타점 38득점을 기록 중이다.
4회 초 KIA는 1사 2루 상황에서 버나디나의 볼넷으로 1사 1, 2루 , 최형우의 볼넷으로 1사 만루 상황을 만들고 나지완의 희생플라이로 1점 앞서나가기 시작했다. 그 후 5회 초에는 2사 2루 상황에서 이명기의 좌익수 플라이로 이닝이 종료되며 달아나지 못했다.
5회 말 SK는 1사 1, 2루 득점찬스를 맞이하였으나 김성현의 중견수 플라이와 1루 주자 로맥의 포스아웃으로 추격에 실패했다. 그 후 6회 말에는 나주환의 1점 홈런으로 승부를 제자리로 돌려세웠다. 2사 1루 상황에서 한동민의 1타점 2루타로 1점 앞서나가기 시작했고, 김동엽의 1타점 2루타로 점수 차를 벌렸다.
7회 초 KIA는 이범호의 2점 홈런으로 승부를 제자리로 돌려세웠다. 2사 2루 득점찬스를 맞이하였으나 이명기의 2루수 땅볼로 아웃카운트를 헌납해 득점에 실패했다. 그 후 8회 초에는 무사 2루 상황에서 최형우를 고의사구로 걸러 1사 1, 3루 상황이 만들어진 후 서동욱의 1타점 내야안타로 1점 앞서나가기 시
Ruby on Rails - S
bash - "drater05-
bash - "ubuntu@
Search Results
Ruby on Rails - F
Stop
Ruby on Rails
Command:
Runner: Ruby on Rails (default)
CWD
ENV
Cannot render console from 183.98.126.74! Allowed networks: 127.0.0.1, ::1, 127.0.0.0/127.255.255.255
Started GET "/assets/logo_big/KIA.png" for 183.98.126.74 at 2017-07-06 23:22:18 +0000
Cannot render console from 183.98.126.74! Allowed networks: 127.0.0.1, ::1, 127.0.0.0/127.255.255.255
Started GET "/assets/logo_big/SK.png" for 183.98.126.74 at 2017-07-06 23:22:18 +0000
Cannot render console from 183.98.126.74! Allowed networks: 127.0.0.1, ::1, 127.0.0.0/127.255.255.255
Started GET "/assets/logo_big/LG.png" for 183.98.126.74 at 2017-07-06 23:22:18 +0000
Cannot render console from 183.98.126.74! Allowed networks: 127.0.0.1, ::1, 127.0.0.0/127.255.255.255

국내 주식시황 기사

[시황기사] 코스피 9.88포인트 상승, 1992.58포인트 거래 마감 (2016.06.22)

22일 코스피 지수는 전날보다 9.88포인트(0.5%) 올라 1992.58으로 상승 마감했다.

이날 외국인과 기관이 각각 255억 원, 1761억 원어치 주식을 사들였으며, 개인만 나홀로 '팔자'에 나서며 2512억 원어치 매도 우위를 기록했다.

시가총액 상위 종목 중에는 오른 종목이 더 많았는데 한국전력(1.02%), 현대차(1.45%) 삼성전자우(0.69%) 등이 상승세를 보였다.

업종별로는 음식료업이 0.61%, 섬유의복이 0.85%, 화학이 0.6% 상승했다.

국내 주식시황 기사 (Cont.)

오늘 코스닥은 전 거래일 대비 0.37포인트, -0.05% 내린 688.18에 마감했다.

기관이 262억 원 매도했으며 개인은 305억 원, 외국인은 74억 원 매수우위를 보였다.

시가총액 상위 종목 중에는 내린 종목이 더 많았는데, 셀트리온(1.16%), 카카오(0.43%) 동서(0.75%)가 상승한 반면, CJ E&M(-0.14%), 메디톡스(-0.82%) 바이로메드(-2.45%) 등은 하락세를 보였다.

업종별로는 유통이 0.31%, 금융이 0.86%, 통신/방송이 0.88% 상승했으며, 건설이 -0.16%, 운송이 -1.63%, 출판매체가 -0.3% 하락했다.

한편 서울 외환시장에서 원 · 달러 환율은 1154.4원에 마감했다.

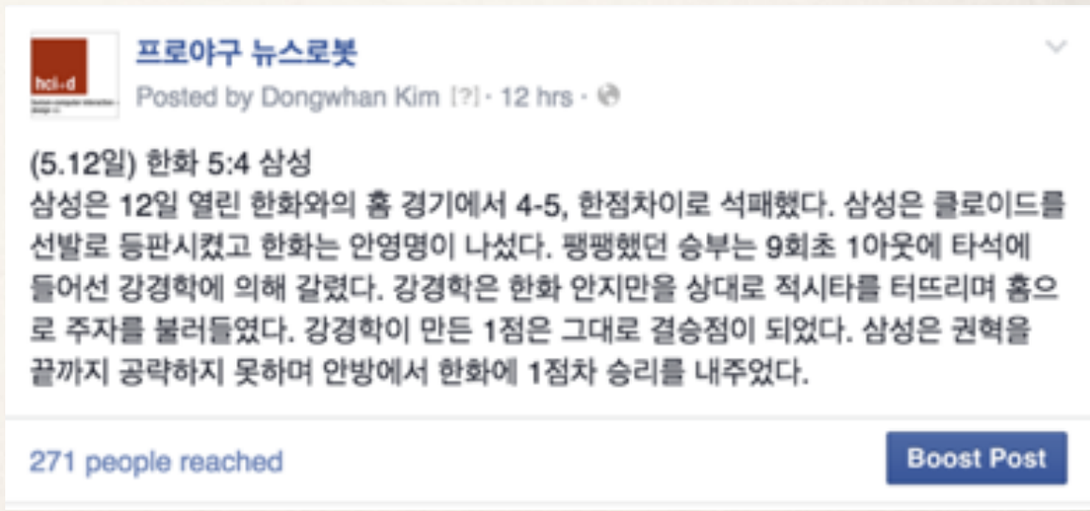
kr-cn-news — python Main.py

```
cond  ['\xec\x9b\x90\xe3\x86\x8d\xeb\x8b\xac\xeb\x9f\xac', 1.6, 1187.7]
lang cn

[시황기사]  (2016-03-15)
15일 코스피 지수는 전날보다 2.3포인트(-0.12%) 내린 1969.97으로 하락 마감했다.

매매주체 별로는 기관이 1293억 원어치를 순매도하면서 지수의 하락세를 홀로 이끌었으며, 개인과 외국인은 각각 95억 원, 611억 원을 순매수했다.

시가총액 상위 종목 중에는 내린 종목이 더 많았는데, 삼성물산(0.34%), 아모레퍼시픽(2.45%)이 상승한 반면, 삼성전자(-0.16%) 등은 하락세를 보였다.

업종별로는 음식료업이 0.2%, 섬유의복이 0.03%, 화학이 0.47% 상승했으며, 종이목재가 -0.11%, 의약품이 -0.42%, 비금속광물이 -0.2% 하락했다.

코스닥은 전날보다 2.29포인트(0.33%) 상승한 693.34포인트로 거래를 마쳤다.

이날 개인과 외국인이 각각 347억 원, 110억 원어치 주식을 사들였으며, 기관만 나홀로 '팔자'에 나서며 399억 원어치 매도 우위를 기록했다.

시가총액 상위 10개 종목 가운데 셀트리온, 카카오가 각각 -0.38%, -1.07%의 하락폭을 보였고, 그 외에 코데즈컴바인(29.92%), 코미팜(0.47%) 등은 상승했다.

업종별로는 운송이 2.15%, 금융이 0.15%, 음식료담배가 0.22% 상승했으며, 건설
```

로봇 저널리즘

로봇 저널리즘

- 기사 작성에 필요한 데이터의 수집에서, 분석, 처리, 기사의 핵심 내용 발견 및 기사 본문 작성, 배포에 이르는 전 과정에 관여하며 자동으로 뉴스 기사를 생성하는 분야
  - 여기서 언급하는 로봇은 소프트웨어 “알고리즘(algorithm)”
- 2015년 부터 한국 프로야구의 모든 경기를 분석해 뉴스서비스 제공
- 파이낸셜뉴스를 통해 2016년 증권시황 기사를 제공
- SBS와 함께 2017년 19대 대선 투/개표율, 득표율 보도 기사 생성



프로야구 뉴스로봇 페이지:
http://www.facebook.com/kbaseballbot

로봇 저널리즘

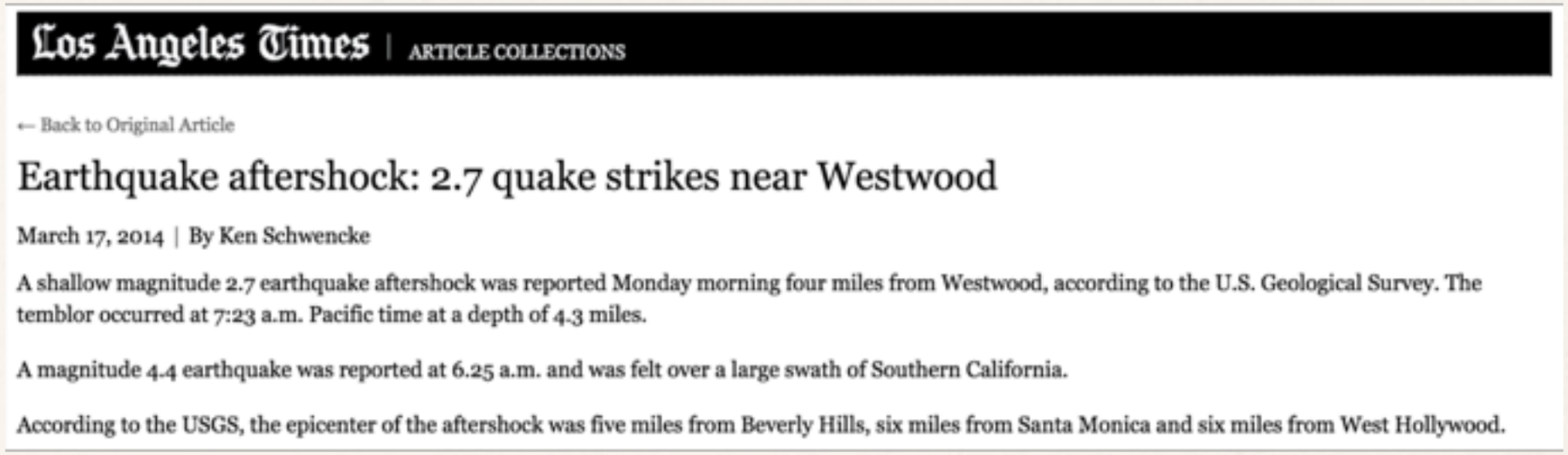
- Northwestern 대학교의 StatsMonkey 프로젝트로 시작 (Templon et al., 2010)
- 이후 ‘Narrative Science’라는 회사 설립.
- ‘Automated Insights’라는 회사와 함께 AP통신, Forbes 등의 매체에 로봇이 작성한 뉴스 기사 공급.
- 국내에서는 서울대학교 hci+d 연구실에서 최초로 한글 기반의 로봇저널리즘 뉴스 기사 생성.

로봇 저널리즘

- 알고리즘은 대용량의 데이터를 빠르고 정확하게 처리할 수 있다는 장점을 가지고 있음
  - 기사에 필요한 데이터를 수집하고 보고서를 작성하는데 걸리는 시간은 불과 0.3초
  - Automated Insight가 2013년 한해 동안 찍어낸 기사의 수는 300만개
- 가까운 시일내에 뉴스봇이 리포터와 편집자보다 빠르게 데이터 속에서 신호(유용한 정보)를 찾아내는 초기 경보 시스템의 역할을 하게될 것
- 재난 상황에서 사람보다 빨리 신호를 감지하여 기사 송고가 가능

로봇 저널리즘

- QuakeBot



Los Angeles Times | ARTICLE COLLECTIONS

← Back to Original Article

Earthquake aftershock: 2.7 quake strikes near Westwood

March 17, 2014 | By Ken Schwencke

A shallow magnitude 2.7 earthquake aftershock was reported Monday morning four miles from Westwood, according to the U.S. Geological Survey. The temblor occurred at 7:23 a.m. Pacific time at a depth of 4.3 miles.

A magnitude 4.4 earthquake was reported at 6.25 a.m. and was felt over a large swath of Southern California.

According to the USGS, the epicenter of the aftershock was five miles from Beverly Hills, six miles from Santa Monica and six miles from West Hollywood.

QuakeBot, 2014/3/17, http://articles.latimes.com/2014/mar/17/news/earthquake-27-quake-strikes-near-westwood-california-rdivor

- Earthquake aftershock: 2.7 quake strikes near Westwood
  - 오전 7시 23분 미국 LA Westwood 지역에서 약 4km 떨어진 곳에서 발생한 지진에 대한 내용을 담은 이 뉴스는 로봇에 의해 쓰여짐
  - LA 타임즈 서버에 자리한 알고리즘이 미국 지질연구소의 이메일을 분석해 지진의 강도에 따라 뉴스 가치를 판단 및 뉴스 발행

로봇 저널리즘

- 이미 영어권의 해외 메이저 언론사를 통해 기자의 손을 거치지 않은, 알고리즘에 의해 자동으로 작성된 뉴스 기사를 다양한 경로를 통해 접하고 있음
- 한국어 환경에서의 로봇저널리즘에 대한 연구가 필요
- 빅데이터 분석, 자연어 처리, 기계 학습 등의 다양한 분야의 다학제적 연구가 필요한 분야

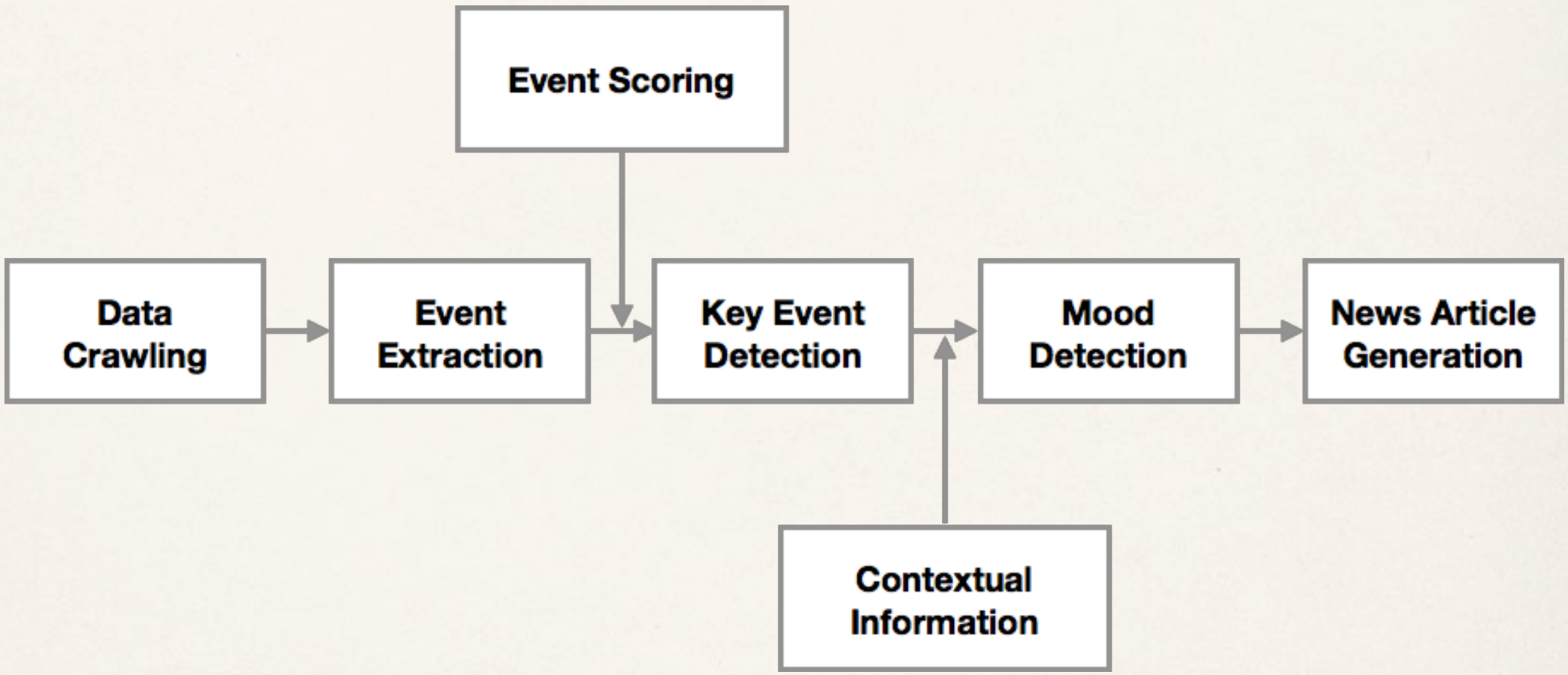
로봇 저널리즘 프레임워크

hci+d lab.

로봇 저널리즘 프레임워크

- 기사 자동 작성을 위한 프레임워크
- 데이터 수집에서 부터 의미 생성, 기사의 작성 까지를 5단계로 구분
  - 데이터 수집 (Data Crawling)
  - 이벤트 추출 (Event Extraction)
  - 중요 이벤트 판단 (Key Event Detection)
  - 무드 판단 (Mood Detection)
  - 뉴스 기사 생성 (News Article Generation)
- 야구기사 작성을 위해 개발되었으나 다른 분야의 기사 또한 작성할 수 있게 개발되어 현재 증시 기사를 만들어 내고 있음.

로봇 저널리즘 프레임워크



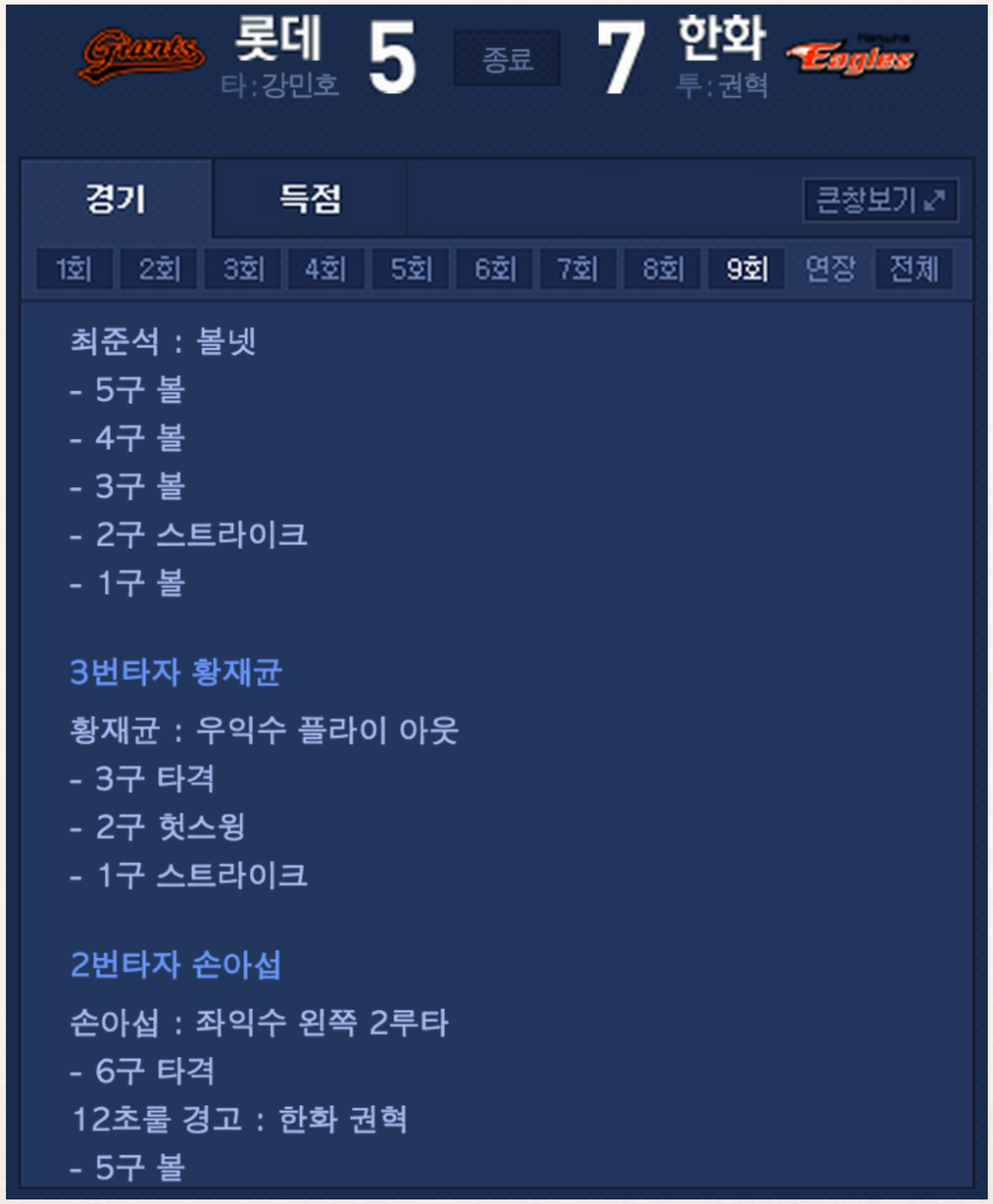
Dongwhan Kim & Joonhwan Lee (2015)

1. Data Crawling

- 데이터를 수집하는 단계
  - API, 공공 데이터, 웹사이트 콘텐츠 및 소셜 미디어 데이터의 크롤링(crawling)
- 데이터에 따라 독자적인 구조를 지닌 비정형 데이터를 분석이 가능한 형태로 변환하는 작업이 수반

1. Data Crawling

- 야구기사의 경우 프로야구 문자중계 데이터 수집



자료제공:
스포츠투아이 (www.sports2i.com)

1. Data Crawling

- Ruby on Rails로 크롤링 소프트웨어 작성
- JSON 형태의 데이터를 수집하여 데이터베이스 구축


```
{"inn":1,"seqno":41,"liveText":"6구 타격","textStyle":1,"btop":0},{"inn":1,"seqno":40,"liveText":"5구 볼","textS
{"inn":1,"seqno":39,"liveText":"4구 파울","textStyle":1,"btop":0},{"inn":1,"seqno":38,"liveText":"3구 스트라이크",
{"inn":1,"seqno":37,"liveText":"2구 볼","textStyle":1,"btop":0},{"inn":1,"seqno":36,"liveText":"1구 스트라이크","
{"inn":1,"seqno":30,"liveText":"3번타자 민병헌","textStyle":8,"btop":0},{"inn":1,"seqno":34,"liveText":"민병헌 :
{"inn":1,"seqno":33,"liveText":"3구 헛스윙","textStyle":1,"btop":0},{"inn":1,"seqno":32,"liveText":"2구 스트라이크
{"inn":1,"seqno":31,"liveText":"1구 스트라이크","textStyle":1,"btop":0},{"inn":1,"seqno":22,"liveText":"2번타자 정
{"inn":1,"seqno":29,"liveText":"2루주자 정진호 : 3루까지 진루","textStyle":14,"btop":0},{"inn":1,"seqno":28,"liveTe
웃)","textStyle":13,"btop":0},{"inn":1,"seqno":27,"liveText":"4구 타격","textStyle":1,"btop":0},{"inn":1,"seqno
진루","textStyle":14,"btop":0},{"inn":1,"seqno":25,"liveText":"3구 스트라이크","textStyle":1,"btop":0},{"inn":1,"
볼","textStyle":1,"btop":0},{"inn":1,"seqno":23,"liveText":"1구 파울","textStyle":1,"btop":0},{"inn":1,"seqno":
호","textStyle":8,"btop":0},{"inn":1,"seqno":21,"liveText":"정진호 : 중견수 앞 1루타","textStyle":13,"btop":0},{"i
격","textStyle":1,"btop":0},{"inn":1,"seqno":0,"liveText":"1회초 넥센 공격 종료","textStyle":0,"btop":1},{"inn":1,
준","textStyle":8,"btop":1},{"inn":1,"seqno":17,"liveText":"유한준 : 3루수 땅볼 아웃 (3루수->1루수 송구아웃)","textStyl
{"inn":1,"seqno":16,"liveText":"2구 타격","textStyle":1,"btop":1},{"inn":1,"seqno":15,"liveText":"1구 볼","textS
{"inn":1,"seqno":7,"liveText":"2번타자 이택근","textStyle":8,"btop":1},{"inn":1,"seqno":13,"liveText":"이택근 : 중
{"inn":1,"seqno":12,"liveText":"5구 타격","textStyle":1,"btop":1},{"inn":1,"seqno":11,"liveText":"4구 스트라이크",
{"inn":1,"seqno":10,"liveText":"3구 볼","textStyle":1,"btop":1},{"inn":1,"seqno":9,"liveText":"2구 볼","textSty
{"inn":1,"seqno":8,"liveText":"1구 볼","textStyle":1,"btop":1},{"inn":1,"seqno":1,"liveText":"1번타자 서건창","te
{"inn":1,"seqno":6,"liveText":"서건창 : 중견수 플라이 아웃 ","textStyle":13,"btop":1},{"inn":1,"seqno":5,"liveText"
{"inn":1,"seqno":4,"liveText":"3구 파울","textStyle":1,"btop":1},{"inn":1,"seqno":3,"liveText":"2구 볼","textSty
{"inn":1,"seqno":2,"liveText":"1구 스트라이크","textStyle":1,"btop":1}],"currentBatter":{"inn":9,"seqno":420,"li
준","textStyle":8,"btop":1},"2":[{"inn":2,"seqno":67,"liveText":"2회말 두산 공격 종료","textStyle":0,"btop":0},{"i
환","textStyle":8,"btop":0},{"inn":2,"seqno":93,"liveText":"1루주자 양의지 : 포스아웃 (유격수->2루수 2루 터치아웃)","tex
```

2. Event Extraction

- 의미있는 이벤트를 추출하는 과정
  - 원시(raw) 데이터에서 정보(information)을 추출하는 과정으로, ‘데이터를 해석하는’ 과정
  - 사람은 직간접 및 관찰 경험, 문화적 배경, 무의식적 직관, 컨텍스트나 분야에 따른 판단을 내림 (Aamodt & Nygård, 1995)
  - 알고리즘도 도메인에 따라 데이터를 해석하는 ‘규칙(rule)’이 필요하고 이에 따라 원시 데이터를 해석

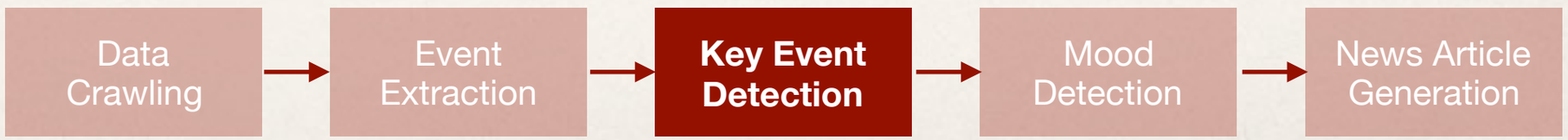
2. Event Extraction

- 타석 별 요약 정보를 이벤트로 추출
  - 루비 언어로 작성된 자연어 처리 소프트웨어를 통해 분석
  - 프로야구 선수들의 데이터베이스를 구축해 자연어 처리 과정에서 참조함 (선수 이름이 중계되면 선수 데이터베이스에서 확인)
- JSON 문자 중계데이터를 분석해 모든 플레이의 상황을 추출하여 “**이벤트**”로 저장

#	원정팀	홈팀	이닝	볼카운트	주자	투수
1	[삼성 0 :	0 넥센]	1회초	(2b:1s)	1아웃 주자 ___루	송신영 > 1번타자 김상수 아웃.
2	[삼성 0 :	0 넥센]	1회초	(2b:3s)	2아웃 주자 ___루	송신영 > 2번타자 우동균 삼진 아웃.
3	[삼성 0 :	0 넥센]	1회초	(0b:0s)	3아웃 주자 ___루	송신영 > 3번타자 나바로 아웃. // 공수교대.
4	[삼성 0 :	0 넥센]	1회말	(0b:1s)	1아웃 주자 ___루	피가로 > 1번타자 고종욱 아웃.
5	[삼성 0 :	0 넥센]	1회말	(3b:1s)	2아웃 주자 ___루	피가로 > 2번타자 박헌도 아웃.

3. Key Event Detection

- 상대적으로 더 중요한 이벤트가 무엇인지 찾아내는 과정
- 중요한 플레이를 찾아내는 통계 기법 사용
  - 경기에서 가장 주목할만한 순간을 찾아내기 위해 야구 통계방법인 세이버메트릭스 알고리즘을 사용
    (증시의 경우 다른 통계 기법 사용)
  - 세이버메트릭스(Sabermetrics)
    - 야구의 다양한 수치를 활용해 평가절하된 투수나 타자를 찾거나, 플레이어가 팀의 승리에 미치는 영향이나 상황 별 기대 득점 등을 통계적으로 산출
    - 머니볼(Lewis, 2004)과 탱고 외(Tango et al., 2007)의 연구로 널리 알려짐

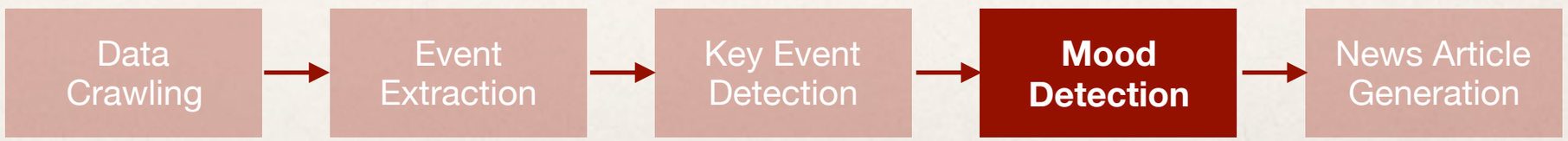


3. Key Event Detection

#	원정팀	홈팀	이닝	볼카운트	주자	투수	중계	기대득점	기대승률	승률변화	중요도	
1	[KIA 0 : 0	넥센]	1회초	(0b:1s)	1아웃	주자 ___루	한현희 > 1번타자 이은총 아웃.	0.54	53.8%	3.8	0.9	Edit
2	[KIA 0 : 0	넥센]	1회초	(1b:2s)	1아웃	주자 1__루	한현희 > 2번타자 강한울 1루타.	0.29	52.6%	-1.2	0.4	Edit
3	[KIA 0 : 0	넥센]	1회초	(2b:1s)	1아웃	주자 12_루	한현희 > 3번타자 필 1루타. 1루주자 2루까지 진루.	0.57	50.5%	-2.1	0.8	Edit
4	[KIA 0 : 0	넥센]	1회초	(3b:3s)	2아웃	주자 12_루	한현희 > 4번타자 최희섭 삼진 아웃.	1.08	50.5%	0.0	1.6	Edit
5	[KIA 1 : 0	넥센]	1회초	(0b:2s)	2아웃	주자 12_루	한현희 > 5번타자 이범호 1루타. 1루주자 2루까지 진루. 2루주자 홈인 +1점추가.	0.58	40.6%	-9.9	1.6	Edit
6	[KIA 4 : 0	넥센]	1회초	(3b:2s)	2아웃	주자 ___루	한현희 > 6번타자 김원섭 홈런 +1점추가. 1루주자 홈인 +1점추가. 2루주자 홈인 +1점추가.	0.58	18.6%	-22.0	1.5	Edit
7	[KIA 4 : 0	넥센]	1회초	(1b:1s)	3아웃	주자 ___루	한현희 > 7번타자 박기남 아웃. // 공수교대.	0.11	15.3%	-3.3	0.2	Edit
8	[KIA 4 : 1	넥센]	1회말	(2b:2s)	0아웃	주자 ___루	서재응 > 1번타자 이택근 홈런 +1점추가.	0.54	26.1%	10.8	0.3	Edit
9	[KIA 4 : 1	넥센]	1회말	(1b:2s)	0아웃	주자 1__루	서재응 > 2번타자 박헌도 1루타.	0.54	29.6%	3.5	0.8	Edit
10	[KIA 4 : 1	넥센]	1회말	(1b:0s)	0아웃	주자 12_루	서재응 > 3번타자 김민성 1루타. 1루주자 2루까지 진루.	0.9	35.3%	5.7	1.4	Edit
⋮												
41	[KIA 5 : 3	넥센]	5회초	(0b:3s)	2아웃	주자 1_3루	한현희 > 5번타자 이범호 삼진 아웃.	1.22	24.4%	0.0	1.5	Edit
42	[KIA 5 : 3	넥센]	5회초	(4b:0s)	2아웃	주자 123루	한현희 > 6번타자 김원섭 볼넷으로 출루. 1루주자 2루까지 진루.	0.52	22.9%	-1.5	1.5	Edit
43	[KIA 5 : 3	넥센]	5회초	(3b:3s)	3아웃	주자 123루	한현희 > 7번타자 박기남 삼진 아웃. // 공수교대.	0.92	21.6%	-1.3	2.2	Edit
44	[KIA 5 : 3	넥센]	5회말	(2b:1s)	0아웃	주자 1__루	박준표 > 3번타자 김민성 1루타.	0.54	33.5%	11.9	4.0	Edit
45	[KIA 5 : 3	넥센]	5회말	(1b:2s)	0아웃	주자 12_루	박준표 > 4번타자 박병호 1루타. 1루주자 2루까지 진루.	0.9	41.8%	8.3	2.1	Edit
46	[KIA 5 : 3	넥센]	5회말	(1b:1s)	1아웃	주자 _23루	박준표 > 5번타자 유한준 아웃. 1루주자 2루까지 진루. 2루주자 3루까지 진루.	1.55	29.8%	-12.0	2.7	Edit
47	[KIA 5 : 5	넥센]	5회말	(1b:2s)	1아웃	주자 _2_루	박준표 > 6번타자 윤석민 2루타. 2루주자 홈인 +1점추가. 3루주자 홈인 +1점추가.	1.28	54.7%	24.9	2.8	Edit
48	[KIA 5 : 5	넥센]	5회말	(3b:2s)	2아웃	주자 _2_루	최영필 > 7번타자 김하성 (투수교체: 최영필) 아웃.	0.8	54.7%	0.0	1.6	Edit
49	[KIA 5 : 6	넥센]	6회말	(1b:3s)	1아웃	주자 ___루	최영필 > 8번타자 이성우 삼진 아웃.	0.54	70.6%	15.9	1.6	Edit
50	[KIA 5 : 6	넥센]	6회말	(1b:1s)	2아웃	주자 ___루	최영필 > 9번타자 오준혁 아웃.	0.29	70.6%	0.0	0.5	Edit
⋮												
58	[KIA 5 : 6	넥센]	7회말	(3b:2s)	0아웃	주자 1__루	최영필 > 4번타자 박병호 1루까지 출루.	0.54	82.0%	7.1	0.4	Edit
59	[KIA 5 : 6	넥센]	7회말	(2b:2s)	1아웃	주자 __3루	최영필 > 5번타자 유한준 (박병호, 도루로 2루까지 진루) 아웃. 2루주자 3루까지 진루.	0.9	78.5%	-3.5	1.2	Edit
60	[KIA 5 : 6	넥센]	7회말	(4b:2s)	1아웃	주자 1_3루	최영필 > 6번타자 윤석민 볼넷으로 출루.	1.06	79.1%	0.6	1.4	Edit
61	[KIA 5 : 7	넥센]	7회말	(2b:1s)	2아웃	주자 1__루	최영필 > 7번타자 김하성 (1루주자 임병욱 교체) 아웃. 3루주자 홈인 +1점추가.	1.22	87.8%	8.7	1.6	Edit
62	[KIA 5 : 7	넥센]	8회말	(1b:1s)	3아웃	주자 ___루	최영필 > 8번타자 박동원 (1루주자 아웃) // 공수교대.	0.24	94.1%	6.3	0.4	Edit

4. Mood Detection

- 뉴스 기사의 분위기를 결정하는 단계
  - 중요 이벤트를 묶어 사건을 바라보는 분위기를 설정
    - 특정 사건이 너무 자주 발생하면 ‘과잉적’ 경향이 보인다고 말할수 있고, 거의 발생하지 않는 일이 발생하면 ‘놀랍게도’라는 수식어가 붙음 → 사건의 패턴
    - 사건의 패턴을 찾는 과정을 통해 기사 전체를 관통하는 일관적인 논조가 생김
    - 예: “NC의 타선, '봇물 터졌다'...LG 상대로 대승”, “KIA와의 아슬아슬한 경기....넥센의 진땀승”
- 단순히 데이터를 수집해 이야기를 만드는게 아니라 독자의 관점에 맞게 개인화된 이야기를 제공할 수 있는 바탕이됨



5. News Article Generation

- 실제 뉴스 기사를 작성
  - 기사 구성 형식에 따른 템플릿을 통한 이야기 생성
    - 중요 이벤트의 데이터를 설명하기 위해 문장의 라이브러리를 구축
    - 무드에 따라 추출된 중요 이벤트르 설명할 수 있는 적절한 문장을 선택해 이야기를 만들어냄
    - 문장에서 주어와 목적어를 비워둔 채로 문장과 데이터를 결합하는 형태 (예, “#{@player_team} #{@player_name}은 #{@inning}회 결승점을 만들며 #{@date}일 경기를 승리로 가져갔다.”)
  - 문장에 변화를 주기위해 탄력적으로 문장을 배열해 기사를 작성할 수 있음

가중치 매트릭스를 사용한 증시 기사의 작성

국내 증시 기사

- 주식 관련 기사 작성 소프트웨어 개발
  - 속보성이 중요한 경우, 사람의 손으로 일일히 작성하는 것보다 빠르게 기사를 작성하고 송고할 수 있음
  - 객관적인 정보의 전달이 가능하고, 관련된 정보를 부가적으로 제공함으로서 컨텍스트를 전달할 수 있음
  - 개인의 포트폴리오 중심으로 작성된 개인화된 기사를 제공
  - 여러 언어로 동시에 기사 작성 가능

데이터 수집(Data Crawling)

- 증권사에서 제공하는 데이터 수집
- 자체 서버에 데이터 베이스 구축
  - 코스피 883개, 코스닥 1,113개 종목
  - 1분 단위로 변동량, 거래량 등의 정보를 DB에 축적
  - 업종별, 시가총액 순위, 해외 증시 변동 등 일단위로 축적
  - API작성으로 간단한 코드를 통해 접근할 수 있도록 함

기사 구성

저가매수에 오름세	대형주 상승 및 하락 이틀 째 코스피 상승	거래대금/거래량 이틀 째 감소
국내증시 현황 - 저가매수세 반등	국내증시 현황 - 이틀째 상승	거래대금 및 거래량
국내증시 현황 - 저가매수세 및 외국증시 강세	국내증시 현황 - 포인트 현황	최대 거래 종목
국내증시 현황 - 포인트 현황	국내증시 현황 - 장 중 상승 후 하락	상한가 및 하한가 정보
국내증시 현황 - 외국증시 강세로 국내증시 변화	전문가 의견	매도/매수 현황
국내증시 현황 - 저가매수세 반등	매수/매도현황	시가총액 및 상위종목 정보
매수/매도 현황	업종 별 및 시가총액 상위종목	
시가총액 상위종목	특징주 - 상승주	
전문가 의견	특징주 - 하락주	
환율정보	국내증시 현황 (코스닥)	
	매수/매도 현황(코스닥)	
	코스닥 특징주	
	환율 정보	

- 템플릿 (template)
- 모듈 (module)
- 이벤트 (event)

템플릿(template)

- 전체 기사의 틀
- 주식 시황 기사의 경우 과거 기사를 분석한 결과, 정형화된 템플릿들이 존재
- 로봇은 수집된 이벤트 데이터를 기반으로, 이에 따라 어떤 템플릿으로 작성할 지 결정
- 최소 한 개 이상의 모듈로 구성되어 있음

모듈(module)

- 일반적으로 하나의 문단을 나타내는 단위
- 모듈의 조합에 따라 템플릿을 결정함
- 언제든지 새로운 모듈을 추가하여 템플릿을 더 자연스럽게 업데이트 할 수 있음
- 예: 국내 증시 현황, 주체별 매수/매도 현황

이벤트(event)

- 모듈을 구성하는 개별 사건
  - 하나의 모듈은 다수의 이벤트로 구성
- 다수의 변수들을 통해 도출된 하나의 의미있는 사건
- 예: 연속 n일간 외국인 매매 상승, 전기전자 업계 거래 변동, 시가총액1위 종목의 주가 변동

가중치 매트릭스(weight matrix)

- 각 이벤트 마다 고유한 가중치들을 행렬로 구성하여 이벤트 간 상대적 중요성 판단의 기준을 삼음
- 각각 이벤트에 대한 **고정 가중치**와 **변동 가중치**로 구성
- 두 가지 가중치를 곱한 값을 개별 이벤트의 스코어로 도출
- 스코어의 총합을 활용
  - $S_{total} = \Sigma (W_{static} \cdot W_{dynamic})$
    - s: 해당 기사의 스코어의 합
    - w: 가중치

가중치 매트릭스(weight matrix)

- 프로야구 이벤트의 예
  - 일반적인 상황
    - 1점 홈런 < 2점 홈런 (고정 가중치)
  - 8회초 0:0 상황에서의 1점 홈런

    8회초 7:2로 이기는 팀의 2점 홈런
    - 1점 홈런 > 2점 홈런 (변동 가중치)

고정 가중치(static weight)

- 기사 가치가 높은 이벤트에 높은 가중치를 부여
- 이벤트의 일반적인 중요성을 의미함
  - 예: 코스피 지수 변동 > 외국인 매매량
- 과거의 기사를 통해 알아냄
- 모든 이벤트가 가지고 있는 개별적인 가중치
  - 초기에는 휴리스틱한 방법으로 결과물을 보고 조정하는 방식
  - 기계학습을 통한 가중치 부여

변동 가중치(dynamic weight)

- 고정 가중치만 가지고 기사를 쓰면 빈 칸 채우기 식의 단조로운 기사가 생성됨
- 맥락적 정보를 가지고 고정 가중치에 곱해지는 값
  - 발생한 이벤트는 1.0을 기본으로 관련된 변수값을 반영
    - 수치적 정보가 반영됨
      - 예: 20일 연속 상승은 3일 연속 상승보다 높은 값을 가짐
    - 발생하지 않은 이벤트에는 0으로 둠
  - 기사의 흐름을 결정하는 핵심적인 값

가중치 테이블과 기사 예시 1

	이벤트	고정 가중치	변동 가중치	스코어
코스피지수	변동량	10.0	7.48	**74.83**
	연속일 변동	8.5	0.33	2.81
	특정선 돌파	2.0	0.0	0.0
	최저/최고값 갱신	3.5	0.0	0.0
코스피 외국인	매매량	5.5	1.86	10.23
	연속일 매매	4.0	0.33	1.33
…	…	…	…	…

[시황] **코스피, 55.52p 상승 1934.2 마감** (2015-09-09)
9일 코스피지수가 급등하며 전일보다 55.52포인트(2.96%) 오른 1934.2 포인트로 장을 마쳤다.매매 주체 별로는 기관이 5천382억원어치를 순매수하면서 지수의 상승세를 홀로 이끌었으며, 개인과 외국인 각각 3천886억원, 1천477억원을 매도했다.

…

가중치 테이블과 기사 예시 2

	이벤트	고정 가중치	변동 가중치	스코어
코스피지수	변동량	10.0	1.43	14.33
	연속일 변동	8.5	0.33	2.81
	특정선 돌파	2.0	0.0	0.0
	최저/최고값 갱신	3.5	0.0	0.0
코스피 외국인	매매량	5.5	2.65	13.25
	연속일 매매	4.0	4.0	**16.00**
…	…	…	…	…

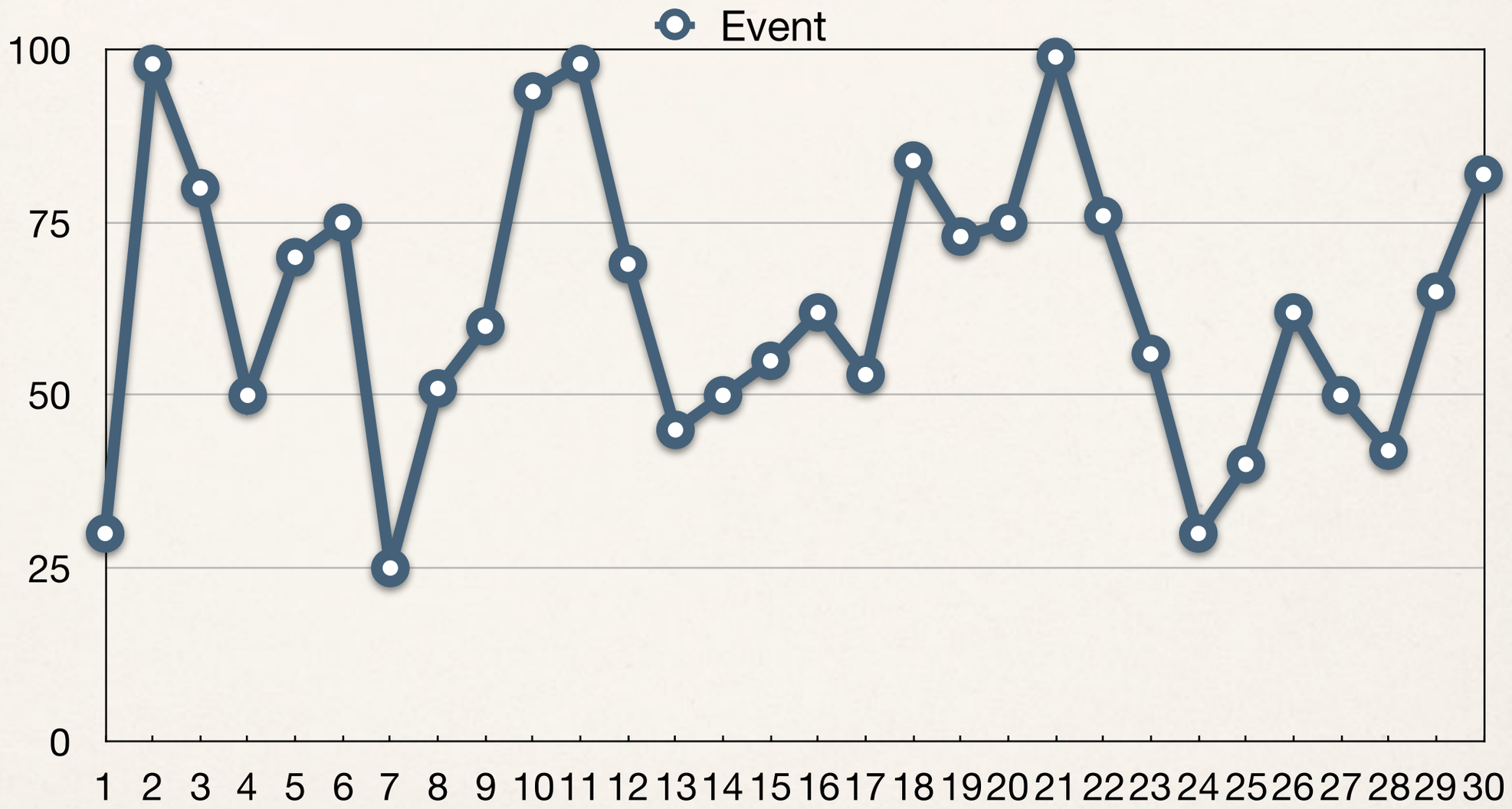
[시황] **외국인 12일째 순매도 코스피 하락 마감** (2015-09-15)
15일 코스피는 전날보다 6.1포인트(0.32%) 상승한 1937.56포인트로 거래를 마쳤다.매매 주체 별로는 기관이 2천111억원어치를 순매수하면서 지수의 상승세를 홀로 이끌었으며, 개인과 외국인 각각 1천341억원, 1천183억원을 매도했다.

…

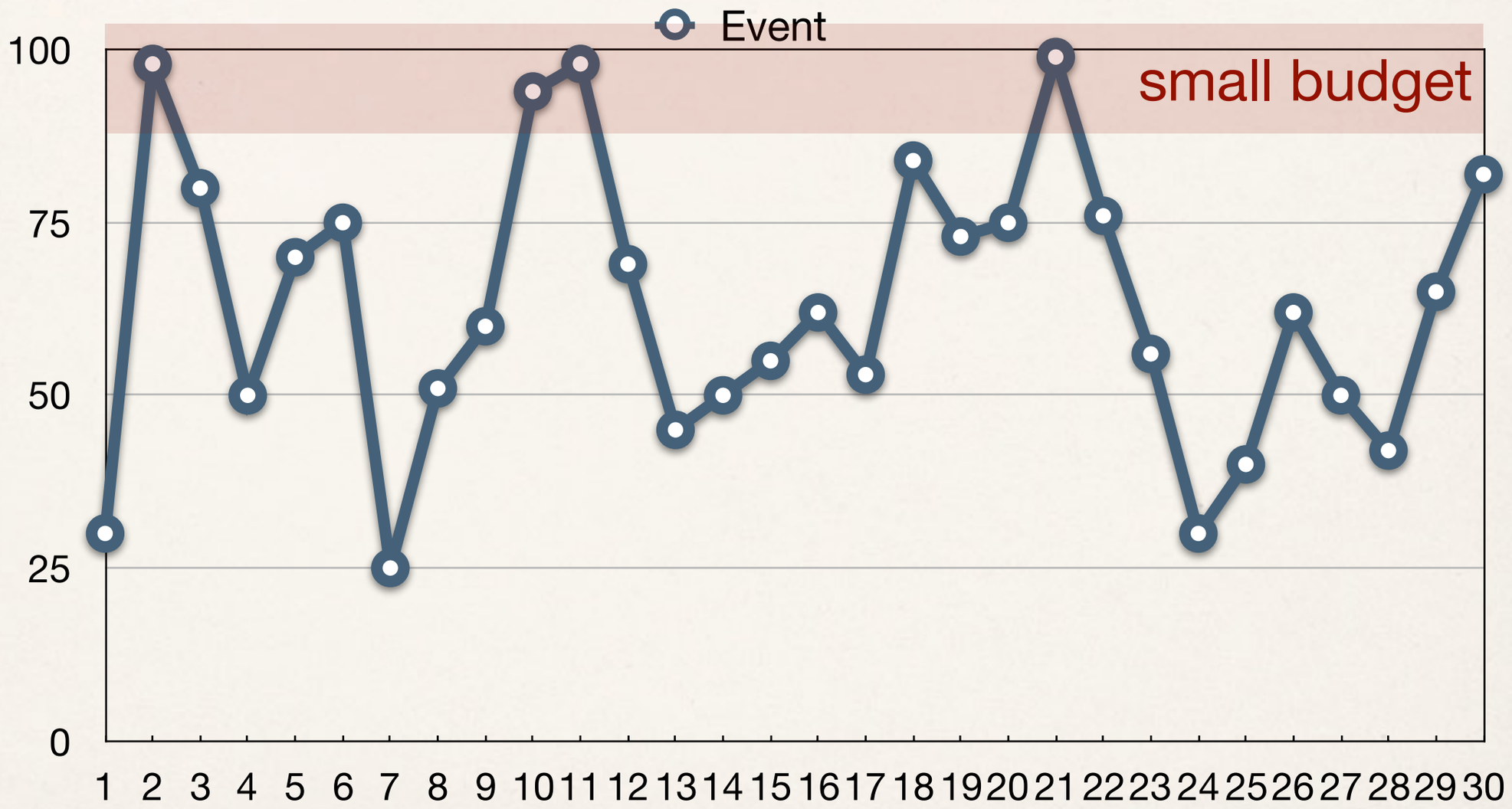
스코어링(scoring) 시스템의 활용

- 기사 제목 선정
  - 가장 높은 스코어를 이용하여 제목을 선정하고 템플릿을 결정
- 기사의 중요도 결정
  - 개별 이벤트 스코어의 합으로 기사문의 중요성을 판단
- 독자의 컨텍스트에 맞는 분량의 기사 출력
  - 개인화
  - 디바이스 최적화

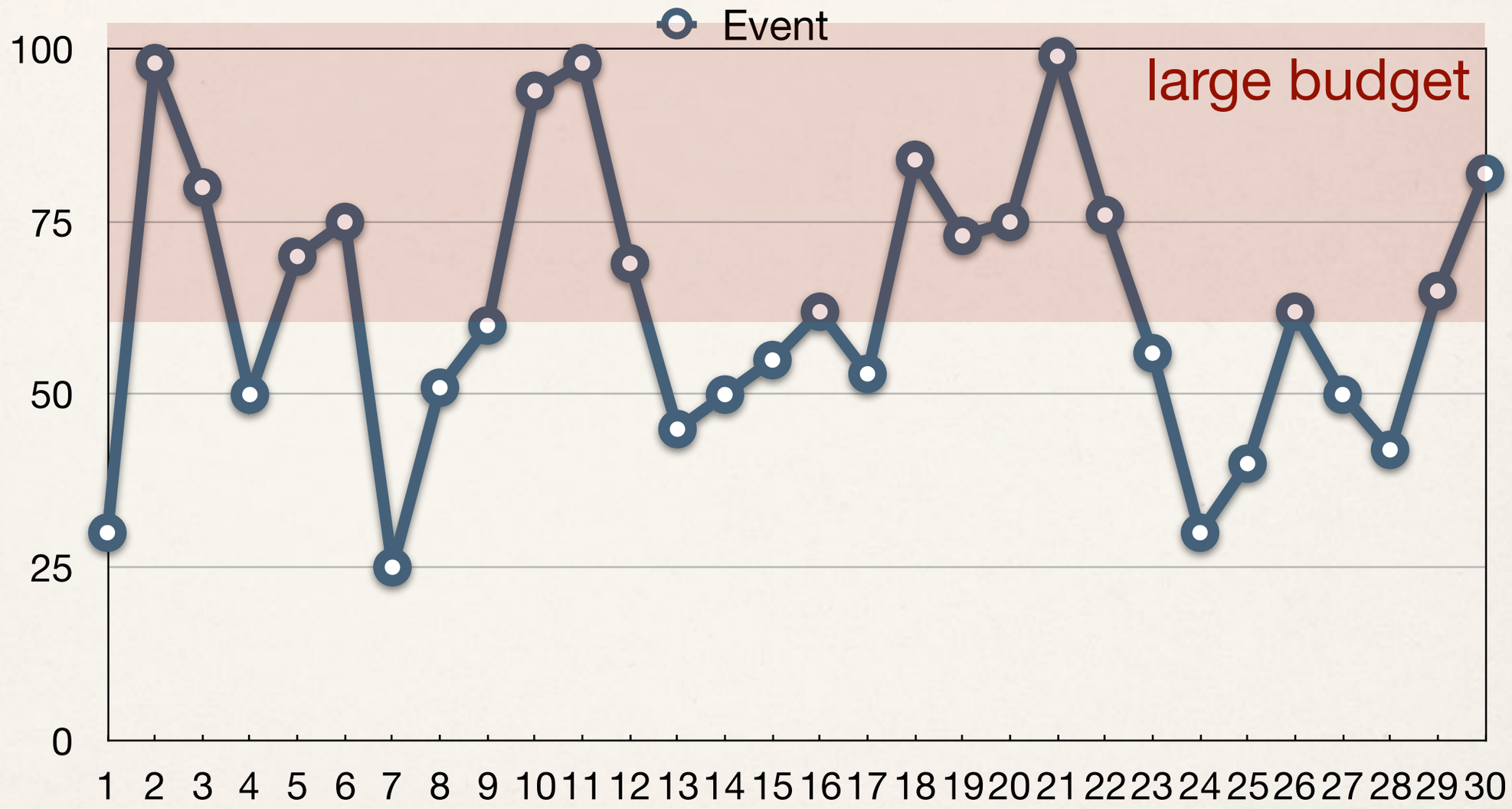
이벤트의 선택


Event
100
75
50
25
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30

이벤트의 선택


Event
small budget
100
75
50
25
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30

이벤트의 선택


Event
large budget
100
75
50
25
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30

컨텍스트에 맞는 기사 생성

- 웨어러블 기기의 보편화, 사물인터넷 발달 배경
- 기사를 보는 컨텍스트에 맞게 실시간으로 기사의 생성
- 사용자의 상황에 따라 다른 길이의 기사를 소비함

나리: 인공지능 로봇기자

- 19대 대통령 선거 보도를 위해 개발 (with SBS)
- 투표율, 개표율, 득표율을 실시간/주기적으로 보도
  - 정해진 시간에 전국/지역별 득표율 관련 기사 생성
  - 특정한 이벤트가 발생했을 때 (예: 역전 시) 속보 기사 생성

http://news.sbs.co.kr/news/reporterPage.do?reporterId=NARe

나리: 인공지능 로봇기자

- 대선 당일 11시 부터 익일 개표 완료 시 까지 **총 250여개의 기사 생성** 후 SBS 송고
  - https://www.facebook.com/robotjournalism/
- SBS에서는 이중 **170개를 선택하여 기사로 발행**
  - http://news.sbs.co.kr/news/reporterPage.do?reporterId=NARe

나리 생성 기사 사례

[NARe] 문재인 후보 당선 확정

제19대 대통령선거에서 더불어민주당 문재인 후보가 당선 확정됐습니다.
오전 2시 41분 현재까지 개표작업이 진행 중인 상황에서, 문 후보는 40.1%의 득표율로 1위를 확정했습니다. 문 후보는 현재까지 1천153만973표를 얻어 자유한국당 홍준표 후보가 얻은 718만8906표보다 434만2067표를 앞서고 있습니다.

이번 대선은 사상 초유의 대통령 탄핵 이후 치러진 선거로, 대통령 당선인은 중앙선거관리위원회에서 당선인 확정을 의결한 직후 바로 임기를 시작합니다. 이에 19대 대통령은 역대처럼 '대통령직 인수위원회'가 꾸려지지 않은 채 대통령직을 수행하게 됩니다.

다음은 문 후보의 현재까지 지역별 득표율입니다(%).

△서울 41.4 △대전 42.1 △부산 37.8 △대구 21.0 △인천 40.6 △광주 60.7 △울산 37.5 △경기 41.2 △강원 33.5 △충북 38.4 △충남·세종 39.3 △전북 64.3 △전남 59.3 △경북 21.0 △경남 35.5 △제주 44.4

나리 생성 기사 사례

[NARe] 문재인 후보 당선 확정

제19대 대통령선거에서 더불어민주당 문재인 후보가 당선 확정됐습니다.
오전 2시 41분 현재까지 개표작업이 진행 중인 상황에서, 문 후보는 40.1%의 득표율로 1위를 확정했습니다. 문 후보는 현재까지 1천153만973표를 얻어 자유한국당 홍준표 후보가 얻은 718만8906표보다 434만2067표를 앞서고 있습니다.

남은 표를 홍 후보가 모두 가져가도 승부를 뒤집을 수 없기 때문에 19대 대통령에 문 후보의 당선이 확정됐습니다.

이번 대선은 사상 초유의 대통령 탄핵 이후 치러진 선거로, 문 당선인은 중앙선거관리위원회에서 당선인 확정을 의결한 직후 바로 임기를 시작합니다.

다음은 문 후보의 현재까지 지역별 득표율입니다(%).

△서울 41.4 △대전 42.1 △부산 37.8 △대구 21.0 △인천 40.6 △광주 60.7 △울산 37.5 △경기 41.2 △강원 33.5 △충북 38.4 △충남·세종 39.3 △전북 64.3 △전남 59.3 △경북 21.0 △경남 35.5 △제주 44.4

나리 생성 기사 사례

[NARe] 문재인 후보 당선 확정

제19대 대통령선거에서 더불어민주당 문재인 후보

오전 2시 41분 현재까지 개표작업이 진행 중인 상

표율로 1위를 확정했습니다. 문 후보는 현재까지

당 홍준표 후보가 얻은 718만8906표보다 43

2067표를 앞서고 있습니다.

데이터를 포함한 문단은 그대로 둔 채 문장을 조금 다듬어서 발행 (평균 5-8분 소요)

남은 표를 홍 후보가 모두 가져가도 승부를 뒤집을 수 없기 때문에 19대 대통령에 문 후보의 당선이 확정됐습니다.

이번 대선은 사상 초유의 대통령 탄핵 이후 치러진 선거로, 문 당선인은 중앙선거관리위원회에서 당선인 확정을 의결한 직후 바로 임기를 시작합니다.

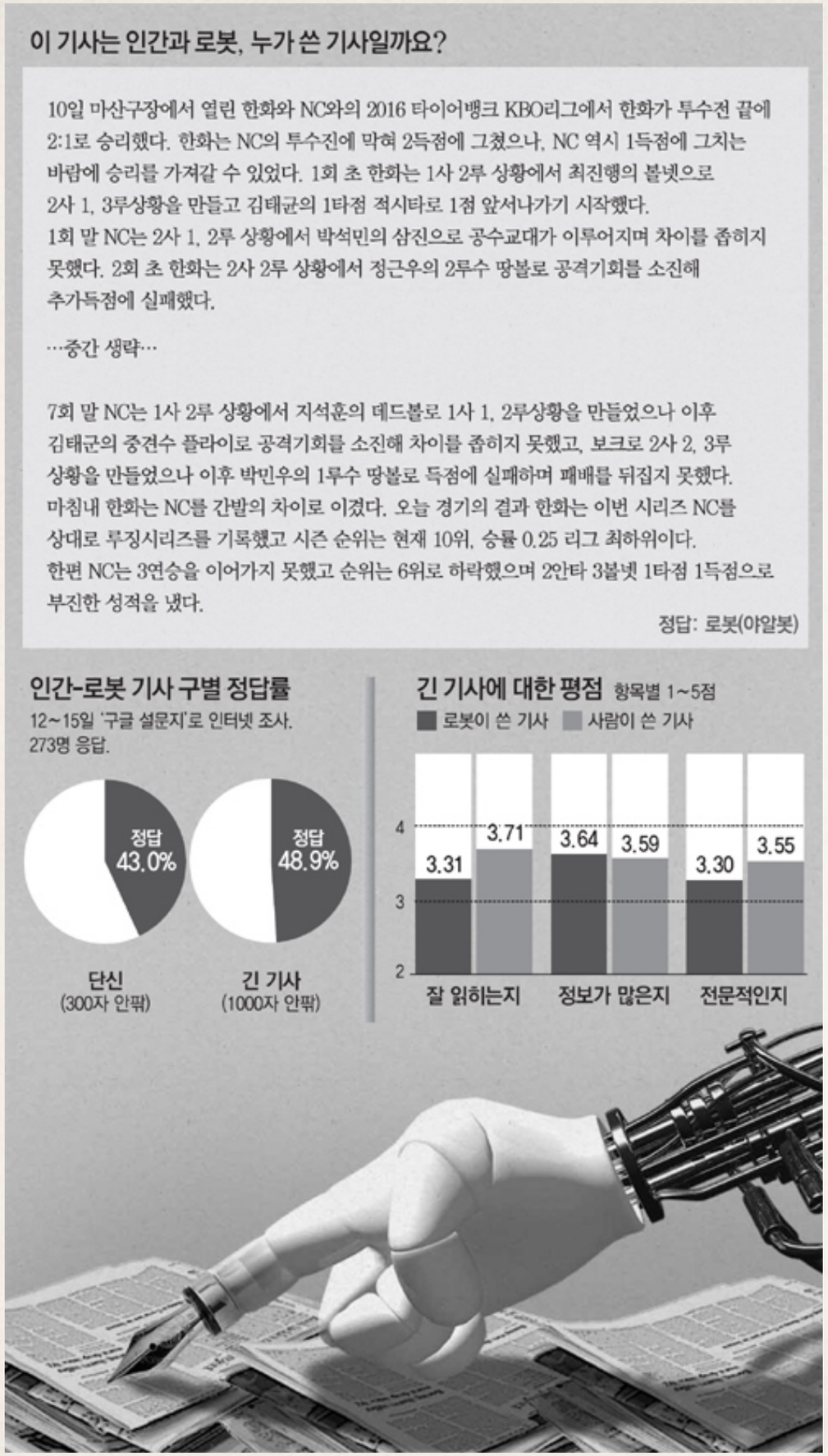
다음은 문 후보의 현재까지 지역별 득표율입니다(%).

△서울 41.4 △대전 42.1 △부산 37.8 △대구 21.0 △인천 40.6 △광주 60.7 △울산 37.5 △경기 41.2 △강원 33.5 △충북 38.4 △충남·세종 39.3 △전북 64.3 △전남 59.3 △경북 21.0 △경남 35.5 △제주 44.4

PINGS :

Algorithm–driven News Generation System for
Personalized and Interactive News Reading Experience

d lab.



- **동아일보에서 12~15일 온라인으로 진행된 퀴즈**
  - 273명 응답
  - 평균 정답률은 45.9% (프로야구 팬이라고 답한 이들의 정답률(46.4%)도 별 차이 없음)
  - 기사의 완성도는 아직 사람이 앞선다고 평가됨. "긴 기사에 대해 '잘 읽히는지' '정보가 많은지' '전문적인지'로 나눠 1~5점 척도로 평가하도록 했는데 사람이 쓴 기사의 평균 점수(3.62)가 로봇(3.42)보다 높았다. 인간 기사가 로봇 기사보다 더 쉽게 읽히고 전문적이라는 평가였다. 다만 정보량에 관한 점수는 로봇이 더 높았다."
  - http://news.donga.com/3/all/20160418/77635201/1

로봇 저널리즘 논의

정보 욕구

- 인지혁명을 만들어 낸 인류의 정보 욕구
  - 단순한 사실관계 정보의 전달
  - 사회적 관계에 대한 정보의 전달
  - 실제로 존재하지 않는 것들에 대한 정보를 전달

- 유발 하라리 <사피엔스>

정보의 폭발적 증가

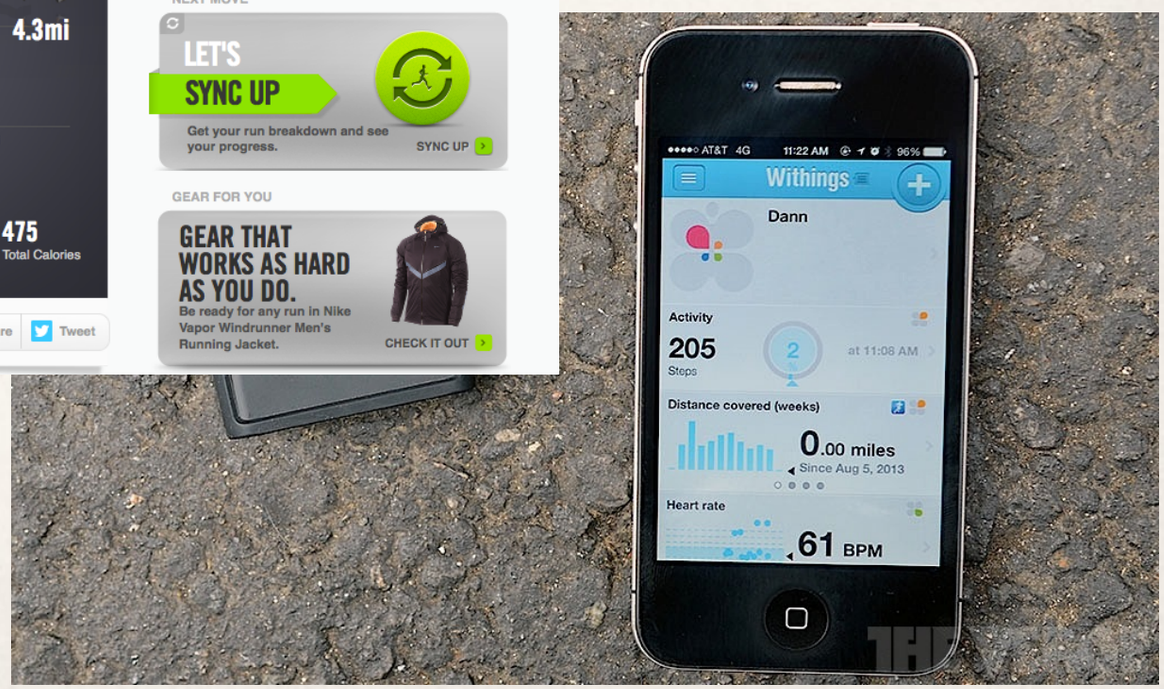
- 오래 전 부터 단순한 정보의 전달은 이루어짐
- 인쇄술/기계의 등장 → 정보의 양적 증가, 확산
- 검증된 정보 필요 → 저널리즘 등장
- 컴퓨터/인터넷 혁명 → 정보가 폭발적으로 증가하고 매우 빠른 속도로 확산 (빅데이터)
  - 너무 많은 정보로 인한 정보 과다 (information overload)
  - 저널리즘의 gatekeeping, agenda setting 축소
  - 검증된 정보 수집의 어려움

It’s All About (Big) Data Analysis

- Large dataset 으로 부터 의미 있는 small dataset 을 선택 → from **data** to **information**
- 사용자에게 의미있는 데이터를 수집하여 적절한 방법으로 사용자에게 전달
- 사용자의 다양한 context 고려하여 데이터 디자인 → situationally appropriate interaction
- 다양한 형태의 전달 방법

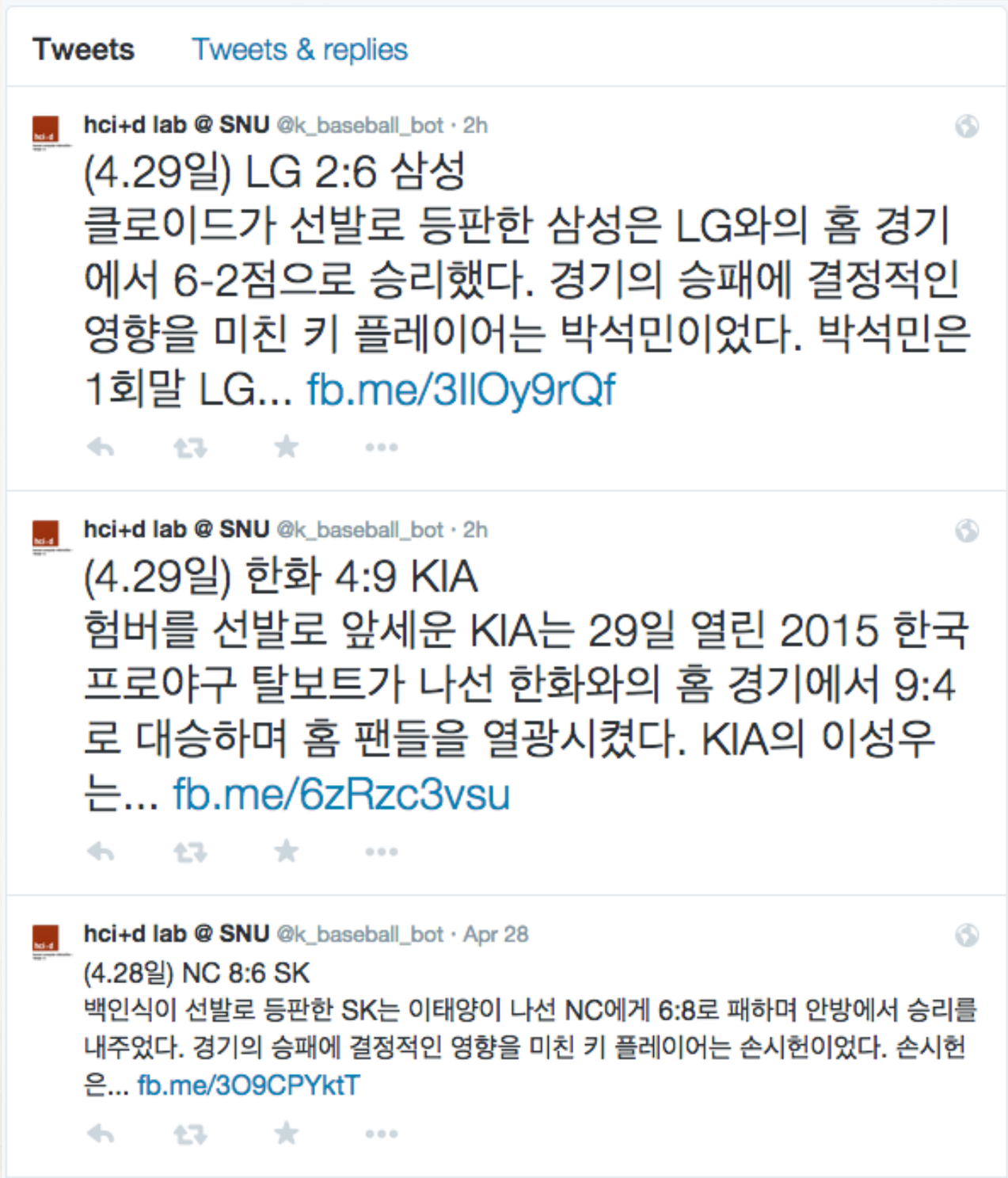
Data를 전달하는 방법

- Visual Form: 데이터의 손쉬운 이해 (external cognition aid)



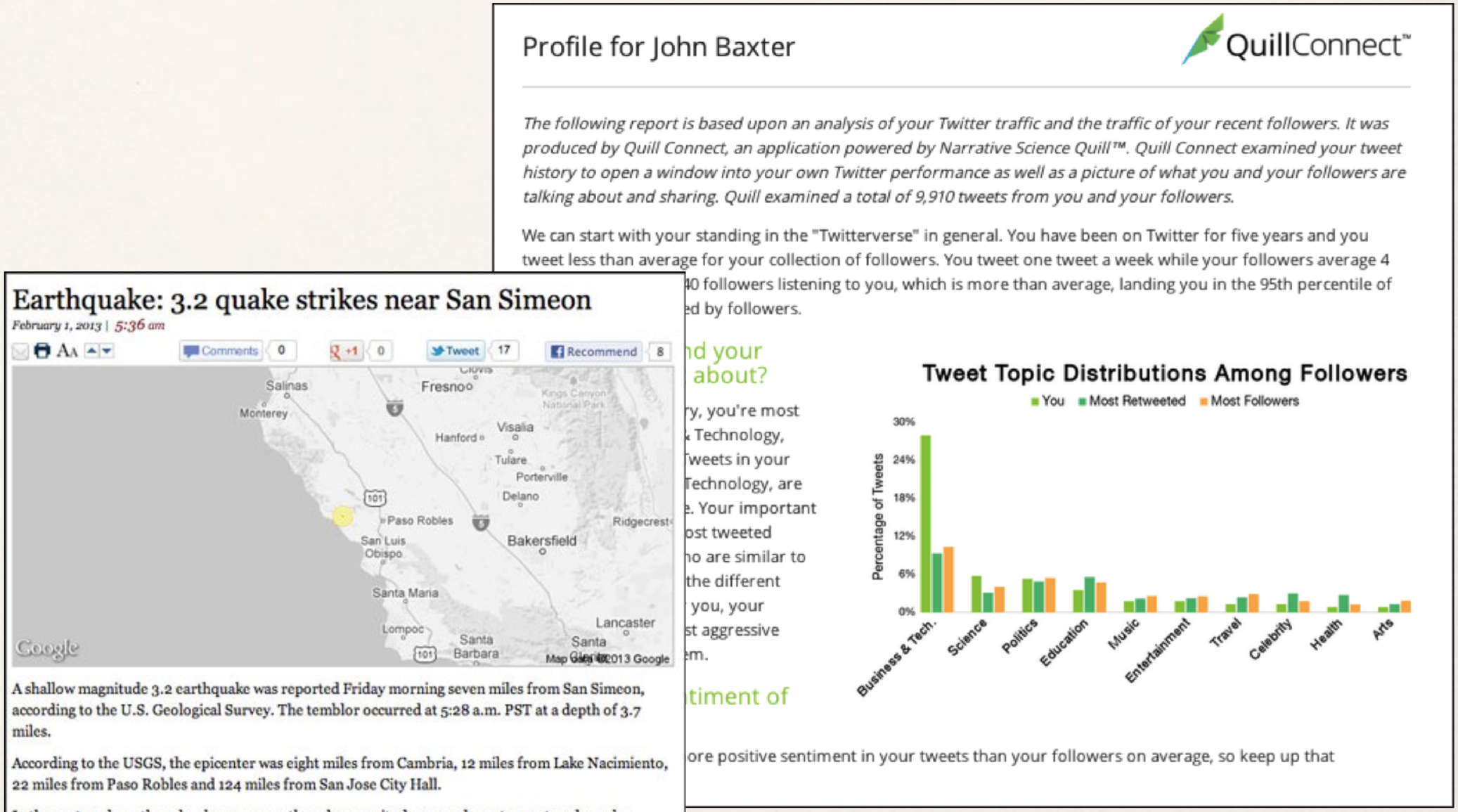
Data를 전달하는 방법

- Narrative Form
  - Robot Journalism
  - 속보성 정보
  - 개인화된 정보
  - 반복되는 정보



Data를 전달하는 방법

- Visual + Narrative Form



Profile for John Baxter

QuillConnect™

The following report is based upon an analysis of your Twitter traffic and the traffic of your recent followers. It was produced by Quill Connect, an application powered by Narrative Science Quill™. Quill Connect examined your tweet history to open a window into your own Twitter performance as well as a picture of what you and your followers are talking about and sharing. Quill examined a total of 9,910 tweets from you and your followers.

We can start with your standing in the "Twitterverse" in general. You have been on Twitter for five years and you tweet less than average for your collection of followers. You tweet one tweet a week while your followers average 4

40 followers listening to you, which is more than average, landing you in the 95th percentile of

ed by followers.

nd your

about?

y, you're most

Technology,

weets in your

Technology, are

. Your important

ost tweeted

no are similar to

the different

you, your

st aggressive

em.

timent of

ore positive sentiment in your tweets than your followers on average, so keep up that

Tweet Topic Distributions Among Followers

You Most Retweeted Most Followers

Percentage of Tweets

30% 24% 18% 12% 6% 0%

Business & Tech. Science Politics Education Music Entertainment Travel Celebrity Health Arts

Earthquake: 3.2 quake strikes near San Simeon

February 1, 2013 | 5:36 am

Comments 0 +1 0 Tweet 17 Recommend 8

Salinas Monterey Fresno Clovis Kings Canyon National Park Visalia Hanford Tulare Porterville Delano Paso Robles Ridgecrest San Luis Obispo Bakersfield Santa Maria Lompoc Santa Barbara Santa Clarita Lancaster 101 5

Google Map data ©2013 Google

A shallow magnitude 3.2 earthquake was reported Friday morning seven miles from San Simeon, according to the U.S. Geological Survey. The temblor occurred at 5:28 a.m. PST at a depth of 3.7 miles.

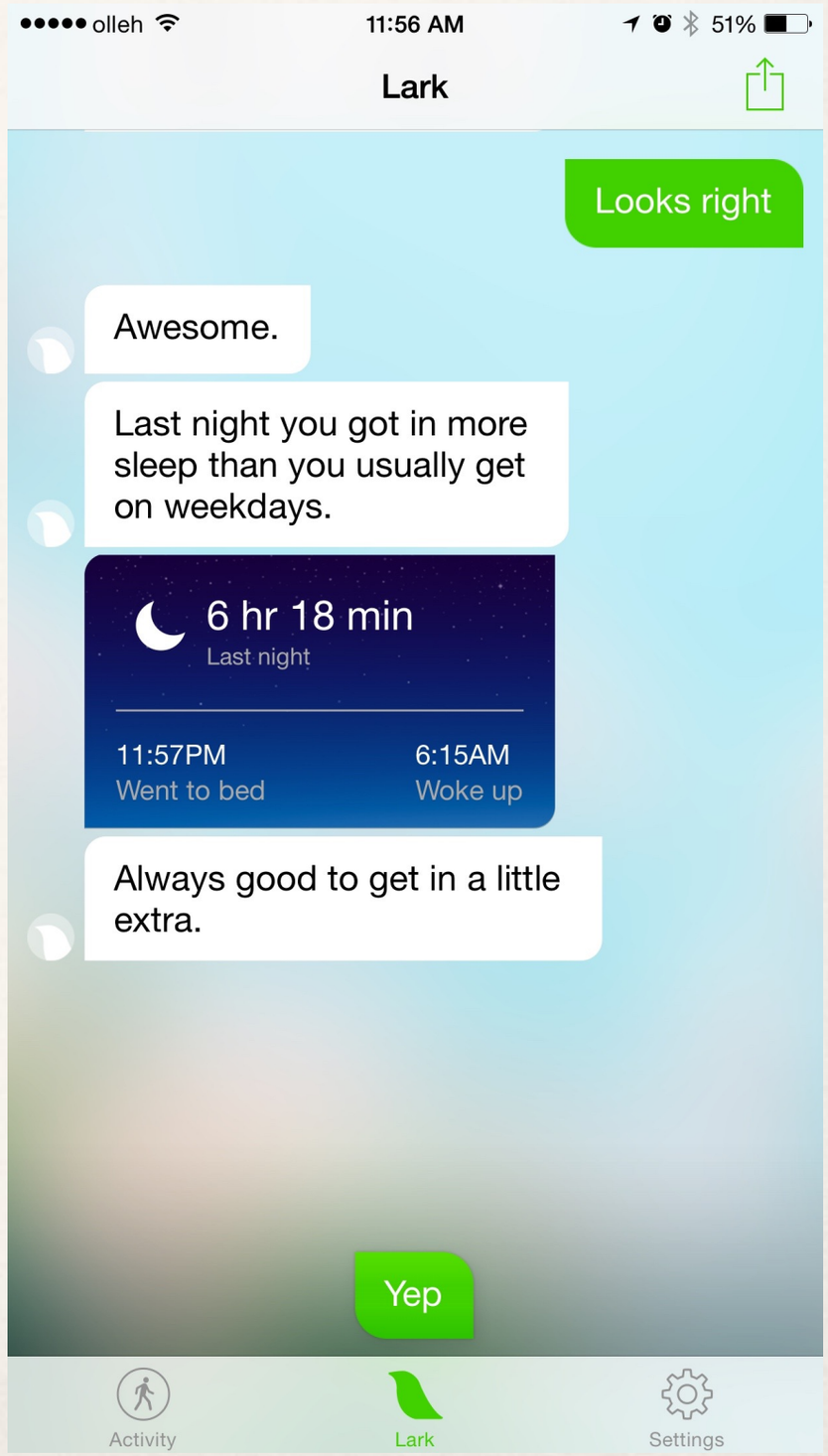
According to the USGS, the epicenter was eight miles from Cambria, 12 miles from Lake Nacimiento, 22 miles from Paso Robles and 124 miles from San Jose City Hall.

In the past 10 days, there has been one earthquake magnitude 3.0 and greater centered nearby.

Read more about California earthquakes on L.A. Now.

Data를 전달하는 방법

- Conversational Form
  - 개인의 컨텍스트를 고려
  - Narrative form 보다 좀더 친밀한 데이터 커뮤니케이션
    - 개개인에게 가치있는 데이터



로봇 저널리즘 연구의 방향

- 로봇 저널리즘을 전통적인 “저널리즘”의 관점에서 평가하기 보다는 폭발적으로 증가한 정보의 바다에서 개인에게 필요한 정보를 **어떻게 찾아서 어떠한 방법으로** 전달할 것인가에 대한 연구로 바라보아야 → 정보 서비스의 관점

로봇 저널리즘 가치

- 중요한 정보의 정확하고 빠른 전달
  - 스포츠 경기나, 주가 동향, 재난 정보 등 빨리 소식을 전하는게 중요한 경우, 사람의 손으로 일일히 작성하는 것보다 빠르게 기사를 작성하고 송고할 수 있음
  - 단순히 빠른 정보 뿐만 아니라 정확하고 객관적인 정보의 전달이 가능하고, 관련된 정보를 부가적으로 제공함으로서 컨텍스트를 전달할 수 있음

로봇 저널리즘 가치

- 개인적인 니즈를 고려한 정보 전달
  - 기존 미디어에서의 기사는 중립적이어야 하지만, 로봇은 독자의 개인적인 취향이나 개인의 포트폴리오, 건강 정보 등을 모두 고려해 개인 맞춤형 기사를 제공하는 것이 가능함
  - 개인의 평범한 일상 속에서 이야기거리를 찾아내 알려주는 일은 로봇 저널리즘이 빅데이터가 아닌 스몰 데이터의 관점에서도 유용하게 사용될 수 있는 가능성을 제시

로봇 저널리즘 가치

- 단순한 일과의 효율적인 대체 수단
  - 데이터를 기계적으로 분석하여 레포트를 반복적으로 작성해야 하는 등의 일은 굳이 사람이 개입되지 않아도 충분히 로봇 저널리즘을 통해 해결 가능
- 데이터를 불러들이는 알고리즘의 강화, 뉴스 가치를 판단하는 기준(통계, 시각화)의 다양화, 다양한 문장을 지닌 템플릿의 작성과 같이 다양한 분야의 전문가들의 노력이 필요한 분야

로봇 저널리즘의 가능성과 한계

- 알고리즘의 편향성이 미치는 영향은?
  - 구글 검색 알고리즘, 아마존 추천 알고리즘 등이 실제로 어떻게 동작하고 어떤 정보를 필터링하는지에 대해 투명하게 알기 힘듦 (필터 버블에 대한 위험)
  - 페이스북 타임라인에 보이는 친구들의 게시물이 필터링되고 있다는 사실에 대한 인지가 페이스북 알고리즘에 대한 (부정적) 의견과 페이스북 이용 형태를 바꿈 (Rader & Gray, 2015)

방송 패러다임의 변화

방송 패러다임의 변화

- 방송도 최근 들어 VOD 나 개인미디어의 소비가 눈에 띄게 증가함
- 신문기사와 마찬가지로 방송 분야도 개인화된 서비스의 등장이 예고되고 있음
- 로봇 저널리즘은 방송 콘텐츠의 개인화에 기여 가능
  - 개인화된 방송 콘텐츠의 생성
  - 개인의 취향과 목적에 맞게 큐레이션된 방송 스케줄

SF 0
LAD 0
1ST
1-1
PITCHES 2
0 OUT
SFGIANTS.COM
SFGIANTS.COM
@SFGIANTS
Majestic
ALWAYS GAME.
KERSHAW
22
GAMESPOT

Automatic Image Annotation

http://googleresearch.blogspot.kr/2014/11/a-picture-is-worth-thousand-coherent.html

Describes without errors	Describes with minor errors	Somewhat related to the image	Unrelated to the image

A person riding a motorcycle on a dirt road.

Two dogs play in the grass.

A skateboarder does a trick on a ramp.

A dog is jumping to catch a frisbee.

A group of young people playing a game of frisbee.

Two hockey players are fighting over the puck.

A little girl in a pink hat is blowing bubbles.

A refrigerator filled with lots of food and drinks.

A herd of elephants walking across a dry grass field.

A close up of a cat laying on a couch.

A red motorcycle parked on the side of the road.

A yellow school bus parked in a parking lot.

방송의 미래

- 날씨, 스포츠, 증시 관련 정보의 방송 콘텐츠 생산 가능
- 지역 정보 또는 개인 관심사에 따른 사회분야 뉴스에 적용 가능
- 알고리즘 기반의 뉴스 편집 시스템을 이용한 개인화된 뉴스 큐레이션 서비스 등장

→ 방송 미디어에 알고리즘을 어떻게 활용할 수 있을지에 대한 고민이 필요한 시점

텍스트를 넘어...

SNU 카드뉴스봇

관심정보 설정하기

설정 비우기

설정 완료

일시

연도는 2018년 월은 6월입니다.

일시는 6월 13일 오전 9시 00분 00초부터 14일 오후 12시 00분 00초(정오)까지 설정가능합니다.

일(13~14)	시(0~23)	분(0~59)	초(0~59)
14	2	0	0

관심 지역

현재 설정된 관심 지역 목록

- 서울특별시 전체 ✖

시도 선택

관심지역 추가하기

관심 선거

- [ ] 국회의원선거
- [ ] 시·도지사선거
- [x] 구·시·군의장선거
- [ ] 교육감선거

관심 후보

현재 설정된 관심 후보 목록

후보검색

관심 정당

- [ ] 더불어민주당
- [ ] 자유한국당
- [ ] 바른미래당
- [ ] 민주평화당
- [ ] 정의당
- [ ] 민중당
- [ ] 대한애국당

Copyright © 2018 hci+d lab. & HCC Lab.
크롬, 사파리, 엣지 및 인터넷 익스플로러 9 이상에서 잘 보입니다.

SNU 카드뉴스봇

내가 만든 뉴스

서울특별시 선거 득표율 현황

#서울시 #구·시·군의 장선거 #06월14일02시00분_기준

11월 06일 13시 23분

업데이트

서울특별시 선거 득표율 현황

#서울시 #구·시·군의 장선거 #06월14일02시00분_기준

11월 04일 13시 09분

업데이트

서울특별시 선거 득표율 현황

#서울시 #구·시·군의 장선거 #06월14일02시00분_기준

11월 02일 21시 27분

업데이트

서울특별시 선거 득표율 현황

#서울시 #국회의원선거 #06월14일09시00분_기준

11월 01일 22시 29분

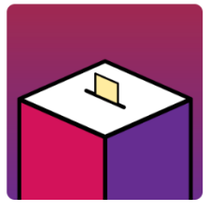
업데이트

서울특별시 선거 득표율 현황

#서울시 #국회의원선거 #06월14일09시00분_기준

11월 01일 22시 27분

업데이트



현재 국회의원선거 득표율 현황

#국회의원선거 #06월14일12시00분_기준

11월 01일 22시 26분

업데이트

현재 국회의원선거 득표율 현황

#국회의원선거 #06월14일12시00분_기준

11월 01일 22시 25분

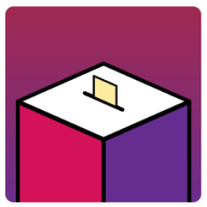
업데이트

현재 국회의원선거 득표율 현황

#국회의원선거 #06월14일12시00분_기준

11월 01일 22시 24분

업데이트



서울특별시 선거 득표율 현황

#서울시 #구·시·군의 장선거 #06월14일12시00분_기준

10월 31일 18시 53분

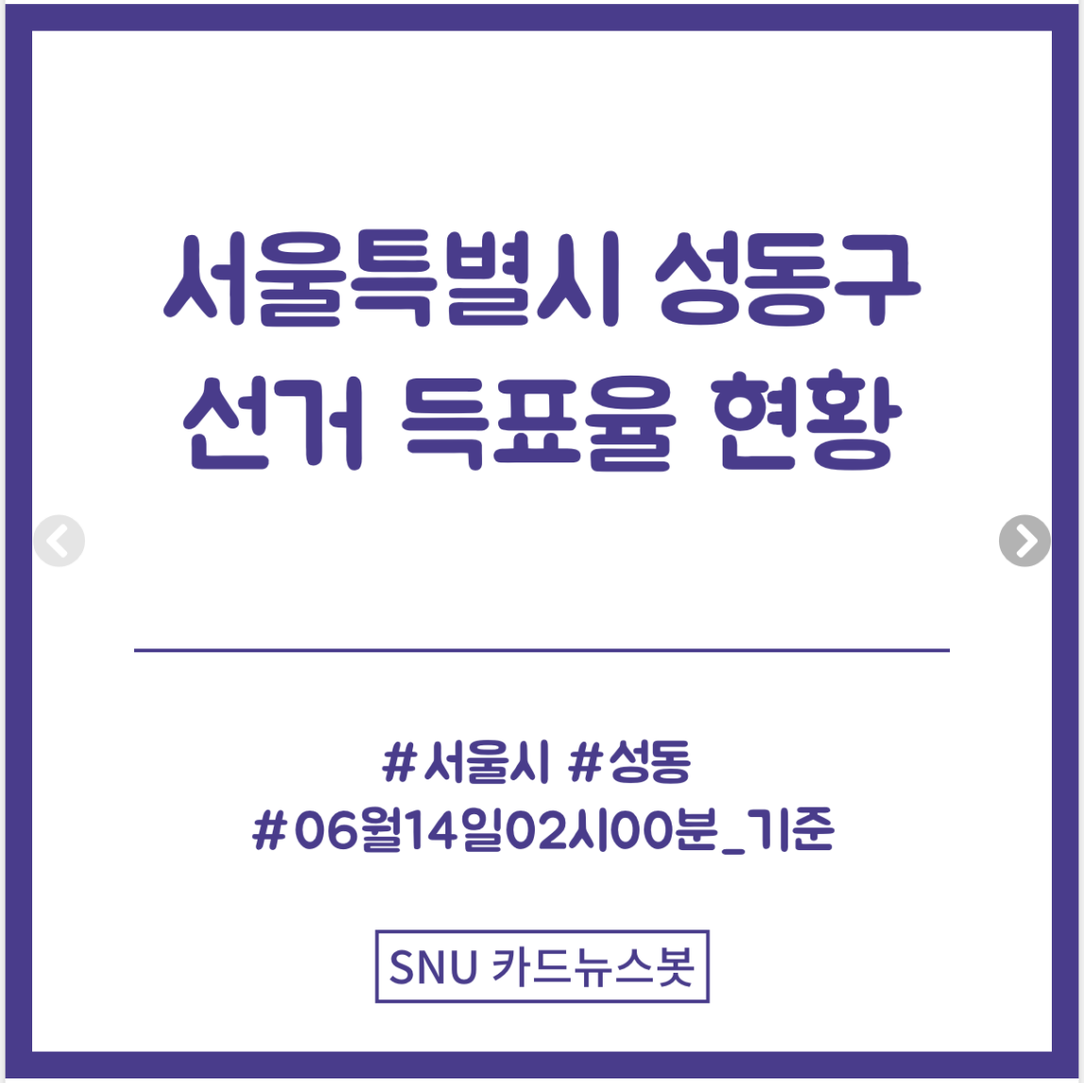
업데이트

Copyright © 2018 hci+d lab. & HCC Lab.
크롬, 사파리, 엣지 및 인터넷 익스플로러 9 이상에서 잘 보입니다.

목록

서울특별시 성동구 선거 득표율 현황

서울시 # 성동 # 06월14일02시00분_기준



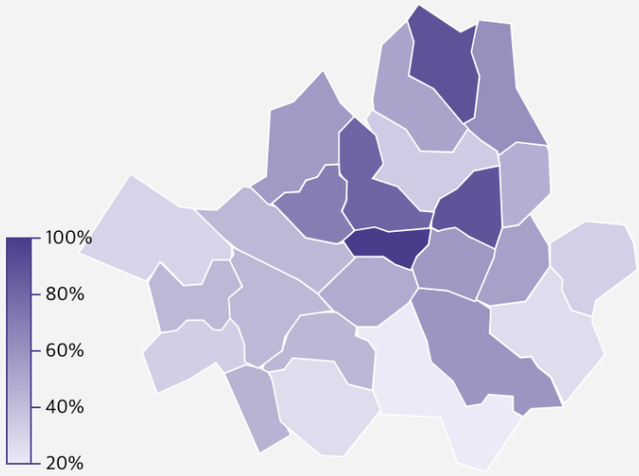
생성 : 2018.11.08 15:24:15

Copyright © 2018 hci+d lab. & HCC Lab.
크롬, 사파리, 엣지 및 인터넷 익스플로러 9 이상에서 잘 보입니다.

목록

서울특별시 성동구 선거 득표율 현황

서울시 # 성동 # 06월14일02시00분_기준



오전 2시 0분 현재 서울특별시 성동구의 개표율은 57.90%로 전국 평균보다 약 8.55% 포인트 낮은 수치이다.

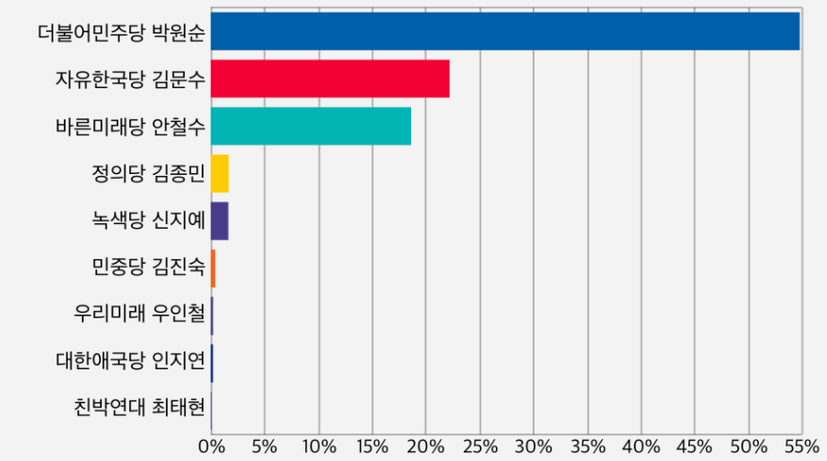
생성 : 2018.11.08 15:24:15

Copyright © 2018 hci+d lab. & HCC Lab.
크롬, 사파리, 엣지 및 인터넷 익스플로러 9 이상에서 잘 보입니다.

목록

서울특별시 성동구 선거 득표율 현황

서울시 # 성동 # 06월14일02시00분_기준



오전 2시 0분 현재 서울특별시 시장 선거에서는 47.47%의 개표율을 보이는 가운데 더불어민주당 박원순 후보가 54.83%의 득표율로 1위, 자유한국당김문수 후보가 22.24%의 득표율로 2위를 차지해 격차가 벌어지고 있다.

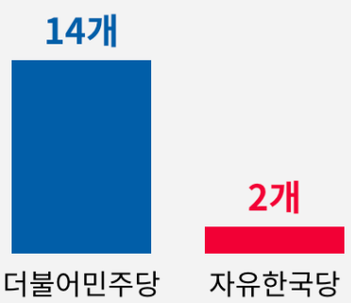
생성 : 2018.11.08 15:24:15

Copyright © 2018 hci+d lab. & HCC Lab.
크롬, 사파리, 엣지 및 인터넷 익스플로러 9 이상에서 잘 보입니다.

목록

서울특별시 성동구 선거 득표율 현황

서울시 # 성동 # 06월14일02시00분_기준



지난 제6회 지방선거에서 새정치민주연합은 시도지사 8석시군구청장 80석 새누리당은 시도지사 7석 시군구청장 117석을 차지했었다. 이번 선거에서는 더불어 민주당이 14개의 시도지사 선거에서 1위를 달리고 있으며 자유한국당은 2개의 시도지사 선거에서 1위를 달리고 있다.

생성 : 2018.11.08 15:24:15

Copyright © 2018 hci+d lab. & HCC Lab.
크롬, 사파리, 엣지 및 인터넷 익스플로러 9 이상에서 잘 보입니다.

Assignment 6: 날씨 기사 생성

- 제출 방법: GitHub (12/1 자정까지)
- 파일 이름: A6-학번-이름(영어로).zip 형식으로 (예: A6-13403-999-jiyoon.zip)

Assignment 6: 날씨 기사 생성

- Robot Journalism Sample Code를 이용하여 날씨 기사 만들기
- 다른 API를 통한 데이터 추가
- 단순한 기사가 아닌 *재미있는* 기사
- 다양한 아이디어를 적용해 볼 것

Questions…?