



Data Science Sentiment Analysis & Fair Eye

2019 한국 정보 올림피아드 코드페어 해커톤 팀: 슬래시 슬립 (//sleep)

2019 평택 한광고등학교 정해솔

2019 동암고등학교 오승현

2019 광주소프트웨어마이스터고등학교 홍하연

2019 멘토 강현욱

시연 사이트 (Running Web Site for Sentiment Analysis & Fair Eye)

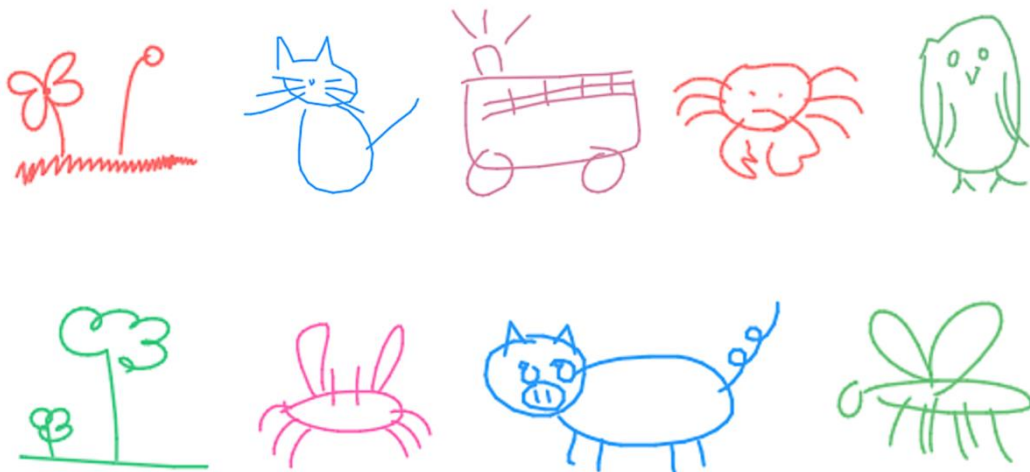
<https://uspressm3-qbkhd.run.goorm.io/>

Presentation

저희 팀은 2019 한국 정보 올림피아드 코드페어 해커톤 준비를 위해 미국 언론에서 다루는 한국 관련 뉴스를 모니터링 할 수 있는 Fair Eye 시스템을 개발해 일반인도 미국 언론이 다루는 한국 정보를 쉽게 찾을 수 있도록 했습니다.

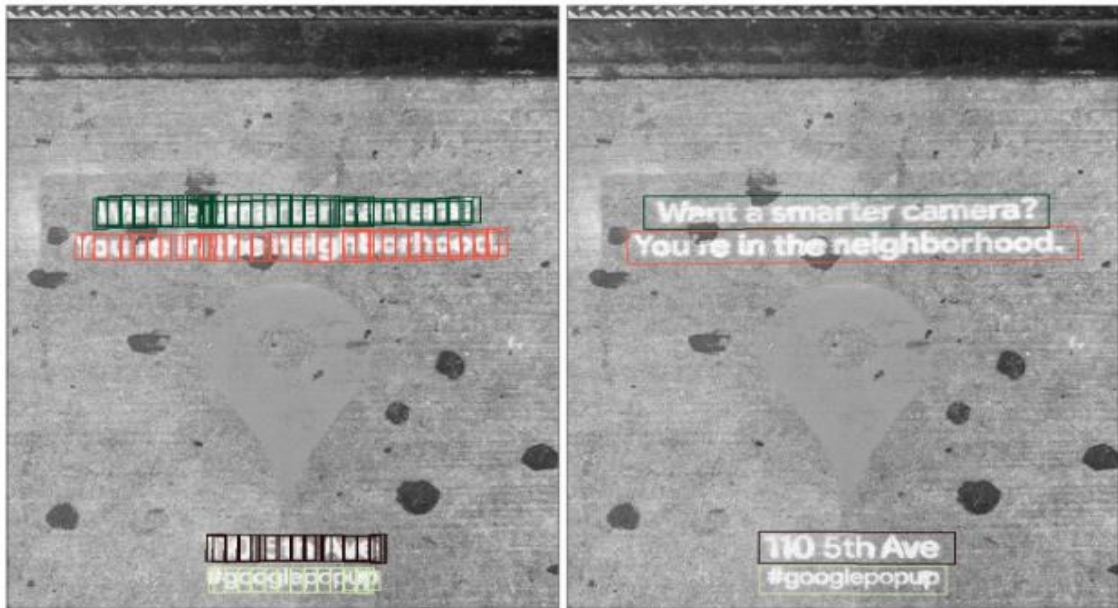
또한 Fair Eye 시스템 개발을 위해 사용한 스크래핑 기술을 이용해서 데이터를 수집하고 분류해 컴퓨터 (AI) 스스로 신문 사설과 기사의 긍정 부정 논조를 파악하는 Data Science Sentiment Analysis System 도 함께 개발했습니다.

최근 구글 AI 연구 방향 중 하나는 어린 아이들 그림을 컴퓨터가 이해하고 아이들이 묘사하는 능력을 컴퓨터에게 학습 시키는 것 입니다. 또한 문서를 자동 인식, 언어를 이해하고 구조를 파악하는 AI 연구를 진행하며 여러가지 서비스를 제공하고 있습니다.



from <https://ai.googleblog.com/2017/04/teaching-machines-to-draw.html>

구글을 선두로 세계 각 기업이 AI 자연어 처리 연구에 막대한 연구비를 쏟는 이유는 전 세계 기업 생존과 직결하는 Big Data 분석을 통한 새로운 마케팅 전략이 AI 자연어 처리 연구에 기반을 두고 있기 때문입니다.



from <https://ai.googleblog.com/2019/09/giving-lens-new-reading-capabilities-in.html>

슬래시 슬립 팀이 개발한 Data Science Sentiment Analysis 시스템은 AI 자연어 처리 방식 일부인 Polarity Detection 파트를 구현했습니다.

```
44 wordlist = content.split()
45 worddictionary = {}
46
47 for word in wordlist:
48     if word in worddictionary:
49         result = re.sub('[^0-9a-zA-Zㄱ-힣]', '', word)
50         # Increase
51         worddictionary[result] += 1
52     else:
53         result = re.sub('[^0-9a-zA-Zㄱ-힣]', '', word)
54         # add to the dictionary
55         worddictionary[result] = 1
56
57 sortedwords = sorted(worddictionary.items(), key=operator.itemgetter(1), reverse=True)
58 print(sortedwords)
59
60 #----- Basic Sentiment Analysis -----#
61
62 #conn = sqlite3.connect("../db.sqlite3")
63 conn=sqlite3.connect("/workspace/USpressM3/db.sqlite3")
64 cur = conn.cursor()
65
66 PositiveNum = 0
67 for word in wordlist:
68     result = re.sub('[^0-9a-zA-Zㄱ-힣]', '', word)
69     #print(result)
70     SQL = "Select * from positivewords where words = " + "'" + result + "'"
71     cur.execute(SQL)
72     rows = cur.fetchall()
73     if rows:
74         PositiveNum += 1
75     print(rows)
76 #print("This article includes:", PositiveNum, "Positive words")
77 print("#####")
78
79 NegativeNum = 0
80 for word in wordlist:
81     result = re.sub('[^0-9a-zA-Zㄱ-힣]', '', word)
82     SQL = "Select * from negativewords where words = " + "'" + result + "'"
```

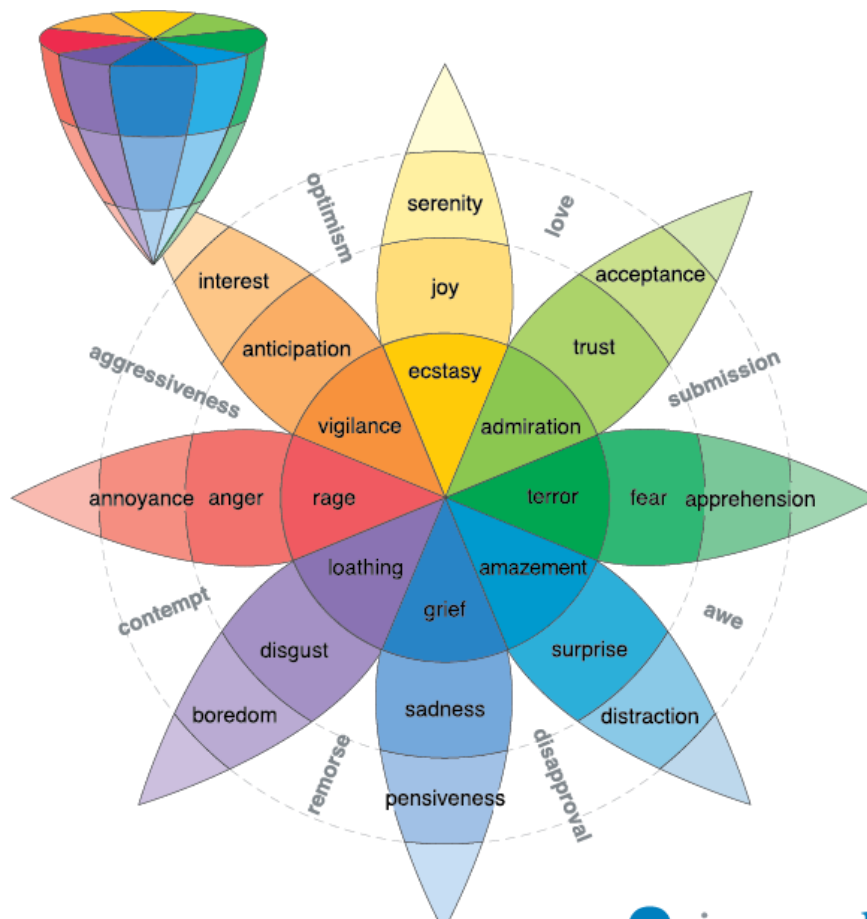
이는 미국 언론을 대상으로 시도하며 모든 소스 코드 공개, 사용된 lexicon 파일을 데이터베이스화 하는 과정 공개, 그리고 관련 알고리즘 뼈대를 알기 쉽게 구현하며 모든 것을 공개하는 것은 저희 팀이 최초라 믿습니다.

지금 보시는 화면은 미국 내 구독자가 가장 많은 USAToday 언론사를 샘플로 한국 관련 뉴스를 실시간 자동으로 가져오게 하고 있습니다.

시연

이를 응용하면 전 미국에 있는 모든 언론사의 한국 관련 기사와 언어적 문제만 해결한다면 전세계 언론을 상대로 한국 관련 기사를 자동으로 탐지하는 시스템을 개발할 수 있습니다.

Robert Plutchik's Emotional Wheel

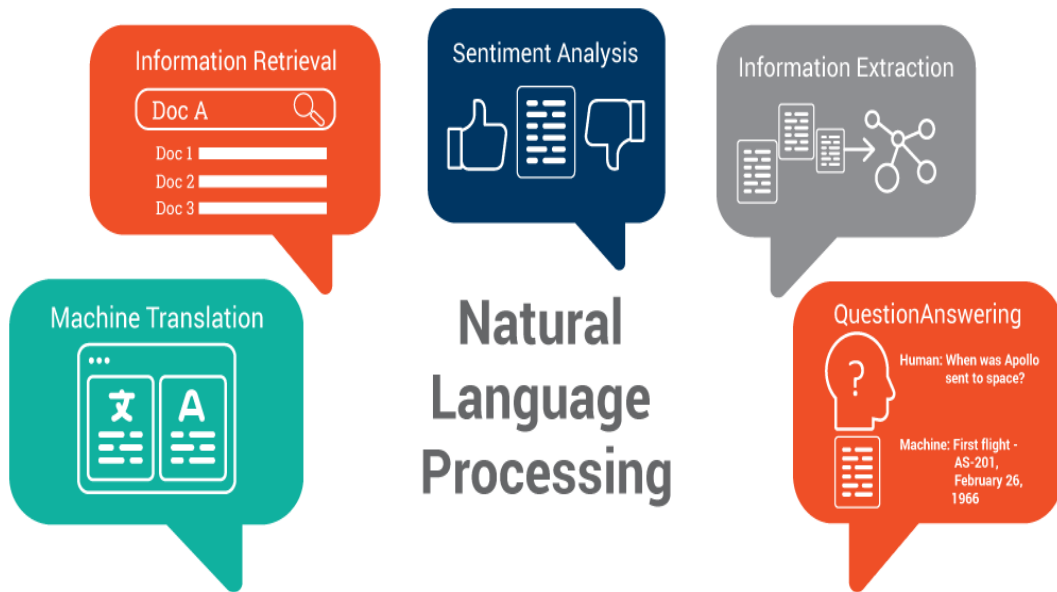


Reprinted on www.6seconds.org by permission of American Scientist, magazine of Sigma Xi, The Scientific Research Society.

 **sixseconds**
www.6seconds.org

이러한 스크래핑 기술을 사용하여 Big Data 분석을 위한 데이터 수집을 쉽게 할 수 있고, 이를 바탕으로 슬래시 슬립 팀은 Robert Plutchik's Emotional Wheel 에서 표현된 여러가지 인간의 감정 중 긍정적인 파트와 부정적인 파트를 AI 스스로 구분하는 Data Science Polarity Detection 시스템을 개발할 수 있었습니다.

지금 보시는 화면은 저희 팀이 만든 Polarity detection 샘플링 테스트 입니다.

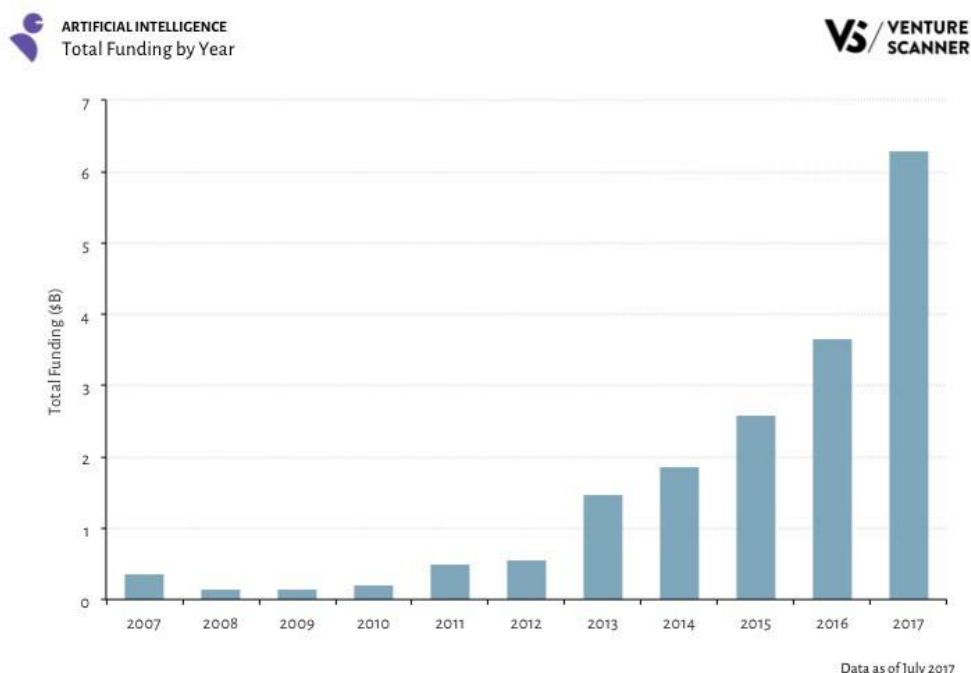


from <https://sigmoidal.io/machine-learning-terminology-explained-top-8-must-know-concepts/>

기사와 사설에 쓰인 단어를 University of Illinois at Chicago Computer Science 에서 제공한 lexicon 긍정 부정 단어와 비교하며 AI 스스로 기사와 사설의 긍정 부정 논조를 파악하고 있습니다. 관련 알고리즘과 작동 방식은 저희 사이트에 자세하게 기술되어 있습니다.

전세계 국가와 기업은 천문학적인 연구비를 투자해 AI 연구에 박차를 가하고 있습니다. 인간이 언어를 배우는 과정이 AI가 자연어 처리 습득 능력을 갖추어 나가는 과정과 비슷하며 정보 가공 능력을 AI 스스로 습득하며 작동할 경우 응용 산업 분야는 방대 합니다.

AI funding has increased over the past few years (AI 관련 기업 투자 증가)



from <https://medium.com/@VentureScanner/artificial-intelligence-funding-trends-q3-2017-4e8a9b80fe53>

세계적인 시장 조사업체 닐슨은 AI 를 이용한 Big Data 분석을 통해 고객에 대한 위험 예측 정보를 제공하고 있고, Google 은 사람들의 관심사를 추적하는 Google Trends 를 통해 각 기업에게 정확한 마케팅 트렌드를 알려주며 방대한 데이터 처리를 위해 AI 를 이용하고 있습니다.

구글은 2018 년 AI 에 관련된 연구를 하는 대학과 비영리 기관에 250 억원이 넘는 연구비를 지원했습니다. 회사 스스로 하는 연구 이외에 전반적인 AI 연구가 미국 전체에서 실행이 되도록 구글은 앞장서고 있습니다.

<https://www.washingtonpost.com/education/2018/10/29/google-offers-money-universities-use-artificial-intelligence-scale-up-research>

미 정부는 새들의 이동경로 파악부터 우주탐사에 이르기까지 산업 전 분야의 AI 연구에 1 조원이 넘는 돈을 순수 연구비로 책정했고 관련 대학과 연구기관에 전폭적인 지지를 보내고 있습니다.

<https://qz.com/1608983/how-much-nasa-dod-and-noaa-are-spending-on-ai-research>

MIT 는 대학 차원에서 1 조원의 넘는 돈을 자체적으로 투자해 AI 대학을 별도로 세우겠다 발표했고 준비 작업에 들어갔습니다.

Artificial Intelligence Oct 15, 2018

MIT has just announced a \$1 billion plan to create a new college for AI



<https://www.technologyreview.com/f/612293/mit-has-just-announced-a-1-billion-plan-to-create-a-new-college-for-ai>

또한 중국은 10 조원이 넘는 돈을 3~4 년 안에 집중적으로 AI 산업에 투자 하겠다 나섰습니다.

FUNDED

China's government AI research has risen 400% since 2007

By Dave Gershgorn • December 13, 2018

<http://theconversation.com/china-is-catching-up-to-the-us-on-artificial-intelligence-research-112119>

아직 한국은 AI 연구가 많이 뒤쳐져 있습니다.

저희 슬래시 슬립팀은 AI Data Science 연구 관련 일부 뼈대를 완성하고 보여줄 수 있다면 관련 연구를 하는 다른 사람들에게 도움이 될 것이라고 판단 했습니다. 그래서 한눈에 이해가 될 수 있는 코드 방식과 심플한 구조를 만들었고 쉽게 다른 연구에 저희 프로그램이 이식될 수 있도록 노력했습니다.

2019 한국 정보 올림피아드 코드페어 해커톤을 통해 저희 팀은 최선을 다 했고 그 결과물이 진심으로 다른 사람들에게 도움이 되길 바랍니다.

감사합니다.

<https://github.com/Dissolve69/DataScienceSentimentAnalysis/>