

발표자 하태윤
심대범
최승환

1.FeedBack

2.개발현황

3.개발 계획

FeedBack

FeedBack

FeedBack



토끼와 거북이는 무엇을 뜻합니까?



FeedBack



성격이 급하고 조바심이 심함.
너무나 **조급**하게 모든 일정을 소화해 내야함.
계획 이행의 **장점**은 있지만, 정신 건강에 좋지 않음.

FeedBack

성격이 **느긋**하고, 사슴 공부법을 추구함.
당장 어떠한 일이 닥치지 않으면 실행에 옮기지 않음.
정신 건강에 **도움**이 되지만, 계획 이행에 있어 **차질이 많음**.





Keyword의 의미는?



FeedBack



어떠한 문장에 대한 **주제**를 나타내는 단어로,
주어에 해당하는 경우가 많다.

EXAMPLE

선거제 개혁이 도대체 왜 문제?



정확한 알고리즘과 개발 현황을 보고해라!

개발 현황

115 5.9

개발 현황



감정 분석 (긍정, 부정)

TFIDF

Text Rank

감정 분석을 통해 각각의 문장 긍정 부정 판단.
TFIDF 를 이용한 Text Rank를 통해 기사에 대한
Keyword 추출.

→ output
해당 Keyword에 대해서 00% 긍정이다, 부정이다

개발 현황

감정 분석 (긍정, 부정)

개발 현황

KoNPLy 형태소 분석 및 품사 태깅 Okt(Open Korean text)를 사용하여 품사 태깅

```
In [5]: from konlpy.tag import Okt
```

```
okt = Okt()
```

```
print(okt.pos(u'지금은 캡스톤 프로젝트 4조 발표다.'))
```

```
[('지금', 'Noun'), ('은', 'Josa'), ('캡', 'Noun'), ('스톤', 'Noun'), ('프로젝트', 'Noun'), ('4조', 'Number'), ('발표', 'Noun'), ('다', 'Josa'), ('.', 'Punctuation')]
```

개발 현황

nltk 라이브러리 사용하여 전처리 작업 실시

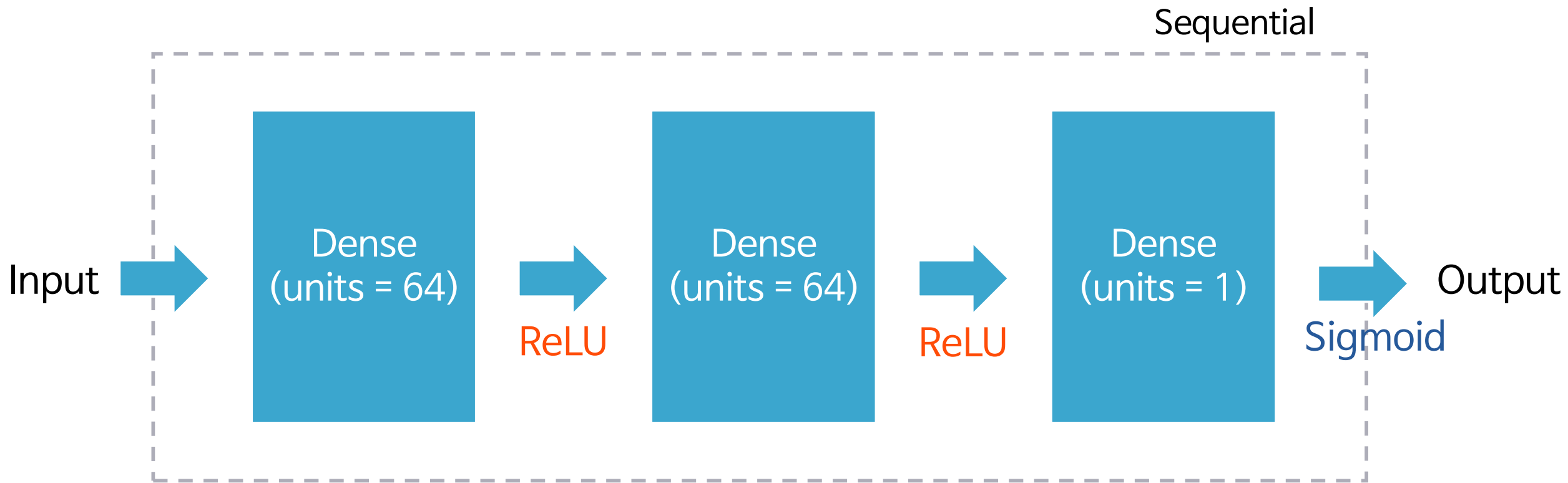
`vocab().most_common` 메소드를 사용하여
데이터에서 많이 사용되는 단어 추출



Count Vectorization 을 사용하여 문서 집합에서 단어 토큰을 생성하고
각 단어 수를 세어 Bag of Words 인코딩한 벡터를 만든다.

개발 현황

IMDB 영화 리뷰 분석 변형 모델 사용



개발 현황

Bianry cross entropy를 손실 함수로 사용

$$BCE(x) = -\frac{1}{N} \sum_{i=1}^N y_i \log(h(x_i; \theta)) + (1 - y_i) \log(1 - h(x_i; \theta)).$$

최적화 알고리즘으로는 RMSProp 통한 경사하강법 사용

$$E[g^2]_t = 0.9E[g^2]_{t-1} + 0.1g_t^2$$
$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{E[g^2]_t + \epsilon}} g_t$$

개발 현황

앞서 말한 알고리즘과 모델을 통해서 미완성 프로그램 구현.
Machine Learning Dataset으로 네이버의 댓글 15000개 사용.

- 댓글과 기사에서 사용되는 용어들은 차이가 큼
- 댓글 TEST 정확도는 약 85%, But 기사 TEST 정확도는 낮음.

```
In [21]: predict("국민 노후자금을 앞세워 경영권을 박탈한 연금사회주의")
```

[국민 노후자금을 앞세워 경영권을 박탈한 연금사회주의]는 69.16% 확률로 부정적인 문장

```
In [20]: predict("사실상 문재인 정권과 계급혁명에 빠진 좌파 운동권들이 죽은 것이나 다름없다")
```

[사실상 문재인 정권과 계급혁명에 빠진 좌파 운동권들이 죽은 것이나 다름없다]는 66.51% 확률로 긍정적인 문장

개발 현황

→ 결론 : DataSet 기사에서 수집 중. 목표치 현재 70% 달성

dataset.txt - 메모장

파일(F) 편집(E) 서식(O) 보기(V) 도움말(H)

법 위에 선 사람들에게서 절 구원해주신 것

진실 밝히기 위해 노력하겠다

사실 이렇게 많은 취재진을 태어나서 처음 본다. 사건사고가 많았는데 범죄의 크기를 무엇이 크고 무엇이 작다고 할 수는 없겠지만...

제가 가장 싫어하는 말 중 '악법도 법'이라는 말이 있다

여기 저를 위해 와주신 분들이 법 위에 선 사람들에게서 저를 구원해주셨다는 생각이 들 정도로 많은 응원과 힘을 얻고 있다

울곧이 걸어왔는데 앞으로도 그 걸음을 여기 와주신 분들과 함께 걸어 나가도록 하겠다

(해당 언론) 기자님, 오셨느냐. 정정보도를 부탁드립니다. 그렇지 않으면 저도 제가 할 수 있는 선에서 법적으로 대응하도록 하겠다

윤 씨의 진실을 향한 투쟁이 외롭지 않도록 국회의원들이 나서서 잘 지켜드리자는 취지로 간담회를 마련했다

윤 씨의 지난주 출판기념회가 석연찮은 이유로 하루 전 취소됐다. 뜻 있는 사람들과 시민의 힘으로 오는 일요일 국회에서 북콘서트를 열겠다

진상 규명이 외압 없이 이뤄지고 핵심 증인 윤 씨가 증인으로서 보호받을 수 있도록 하겠다

진실이 밝혀지도록 노력하겠다

대한민국에서 권력형 범죄를 뿌리 뽑을 수 있도록 저희 국회에서도 최선을 다하겠다

국회가 윤 씨의 방패막이가 되겠다

윤 씨가 겪은 두려움과 외로움의 시간에 대해 국회가 성찰할 것

더 기다릴 사안은 아닌 것으로 보인다

문 대통령은 정상적 절차에 따라 8일 오전 두 후보자에 대한 임명안을 재가할 예정

농지 투기를 한 것은 맞죠. 그래도 의원님들 덕분에 마을에 도로가 생겨서 고마운데 어떻게 길이 개설됐는지 얘기하기 조심스러워요.

고맙긴. 뭐가 고마워? 주민들 이용한 거지.

도로 개설에 노력하겠다

아내가 투기로 사들인 땅에 빌라를 지었다

탁 자문위원에게 민주당 홍보위원장을 맡기로 했다

당 홍보를 강화하기 위해 탁 자문위원의 전문성이 필요하다는 공감대가 있었다

탁 자문위원이 정부·여당의 국정철학을 제대로 이해하고, 이를효과적으로 국민에게 전달할 적임자라는 데에는 이견이 거의 없을 것

민주당에서 아직 홍보위원장직을 제안받은 바 없고, 생각해보지도 않았다

지도부가 총선 승리를 위해 친문(친문재인)과 비문(비문재인)을 아우르는 '원팀' 전략을 고수할 것

당직 인선에서도 그 점을 깊이 고려할 것

인재영입은 당분간 공개적으로 진행할 가능성이 작다

이해한 대표가 영입을 염두에 두고 다양한 분야의 인사들을 만나기 시작했다

법으로 하는 주먹질

그런 일은 없습니다

일본에도 한국과 마찬가지로 명예훼손죄가 있습니다. 무고죄도 있는 것으로 알고 있습니다. 최근 아베 신조 총리는 소위 '학원스캔들' 문제로 아, 그럴 수도 있었군요!

정치적인 논란을 가지고 검찰이나 경찰에 고소, 고발을 하는 일은 거의 없습니다. 의원들 스스로도 고소장을 내지 않고 설사 고소를 한다고 하

○○○의원 ○○ 문제로 피소

총선 때 공지영이 투표한다고 자기 생일 올렸잖아요. 진짜 토할 뻔했어요. 50 먹은 여자가 왜 생일을 올립니까? 공주병은 확실해 보여요

환부 돌려내야

당의 환부를 돌려내고 전진해야 할 때

dataset.txt - 메모장

파일(F) 편집(E) 서식(O) 보기(V) 도움말(H)

공정 검토하겠다

범국민 산?품奘? 추진위원회를 구성하고 규제 완화를 위한 특별법 제정을 건의한다

특히 올림픽이 열리는 평창에서 산악철도를 비롯해 관광사업 활성화로 지역 경제 발전에 크게 기여할 것

산림청이 삼양 땅을 환원받아 수목원을 조성하려는데 삼양이 땅을 내놓으려고 하지 않아 중재한 것이다. 생수와 관련된 발언도 나라나 지방자치·관광농원 등을 하려고 샅었다. 내가 농사를 지은 것은 아니고 공동 매입자가 농사를 지어왔다

땅값을 올리려는 의도로 공약?? 제시하지 않았다

(공약 대상지) 근처에 소유 땅이 있다고 해서 문제라면 지역구에 땅이 있는 의원은 지역 발전을 위해 아무것도 하지 말라는 얘기가 되지 않느냐 이해관계가 걸려 있으면 관련 공약을 제시할 때 재산을 함께 '공개'하는 수준에서 잠재적 충돌 상황을 사전 관리하는 방안도 있다

공약을 추진하되 부동산을 사전에 공개하면 사적 이해관계 반영을 제어하고 모니터링도 할 수 있다

의원에 대한 사후적 감사, 징계보다 사전적 예방과 상담을 하는 것이 (이해충돌) 독립기구의 더 큰 목적이다. 본의 아니게 지탄받지 않도록 미리 자 이해충돌을 잠재적 상태부터 실질적 이해충돌까지 폭넓게 관리하는 이유가 있다

이해충돌 방지 제도가 보호하고자 하는 가치는 공직자가 공정하게 직무를 수행할 것이라는 '신뢰' 그 자체

공약도 실현되고 자신의 이익도 챙기는 잠재적 충돌 상황

이해충돌 관련 입법이 미비한 상태에서 토지와 공약의 연관성을 들여다본 자료는 이해충돌 방지를 입법화할 수 있는 근거가 될 것

국회의원의 경우 일반 공직자보다 포괄적인 업무를 하는데 공약 안에 토지가 있다면 잠재적 이해충돌의 요소가 있다. 다만 규제의 목적과 범위, 토진입과정서 희생된 제주도민에 깊은 유감과 애도

무고하게 희생된 분들께 분명히 사죄

법적인 검토

'제주4·3특별법'의 정신을 존중하며, 진압 과정에서 제주도민들이 희생된 것에 대해 깊은 유감과 애도를 표한다

진상규명을 위한 정부의 노력이 최선을 다해서 적극 동참

계속해서 법적인 검토를 해나가겠다

무고하게 희생된 분들께는 (경찰이) 분명히 사죄를 드려야 한다

진실이 하나하나 밝혀지고 있기에 밝혀진 사실에 따라서 경찰도 인정할 것은 인정할 것

국방부와 경찰청의 사과를 환영한다

애도와 유감을 표명했지만 적극적인 방법의 사과가 아니어서 아쉽다. 진정성 있는 사과의 자세로 다가와주기를 촉구한다

4·3의 완전한 해결이 이념을 극복하고 국민통합으로 가는 길

진상을 완전히 규명하고 배·보상 문제와 트라우마 치유센터 설립 등 제주도민들의 아픈 상처를 치유하기 위한 일에 더욱 힘을 기울이겠다

김의겸 전 대변인, 국민은행서 10억 대출 특혜 의심

김 전 대변인이 국민은행으로부터 상가구입을 위해 대출받은 10억원이 사실과 다르게 조작된 것으로 확인됐다

2층 상가건물에는 10개의 상가가 입주 가능한 것으로 돼 있고, 이에 근거해 월 525만원의 임대료 수입이 가능하다고 계산을 했다

월 10억 대출에 대한 이자는 월 350만원이 넘는데 건축물대장을 확인해보니 4개의 상가만 입주해있어 실제 임대 소득은 300만원도 되지 않는다 전문가들은 이 건물에 10개 상가는 입주가 사실상 불가능하다고 평가하는 상황에서, 대출서류를 조작했거나 국민은행이 재개발을 기대하고 공모 불법적으로 일어난 대출로 검찰은 이 부분에 대한 고발이 없더라도 직권으로 수사에 나서야 한다

김 전 대변인이 지점장과 어떤 거래가 있었는지, 권한을 남용했는지에 초점이 맞춰져야 한다

대출 시기가 지난해 8월로, 당시에는 RTI가 1.5에 충족되지 않더라도 대출 취급이 가능했다

추정상가 갯수를 적용하는 것은 대출한도를 ?捰瘳蛄막? 잡기 위해 사용하는 방식

Text Rank를 통한 주제어 추출

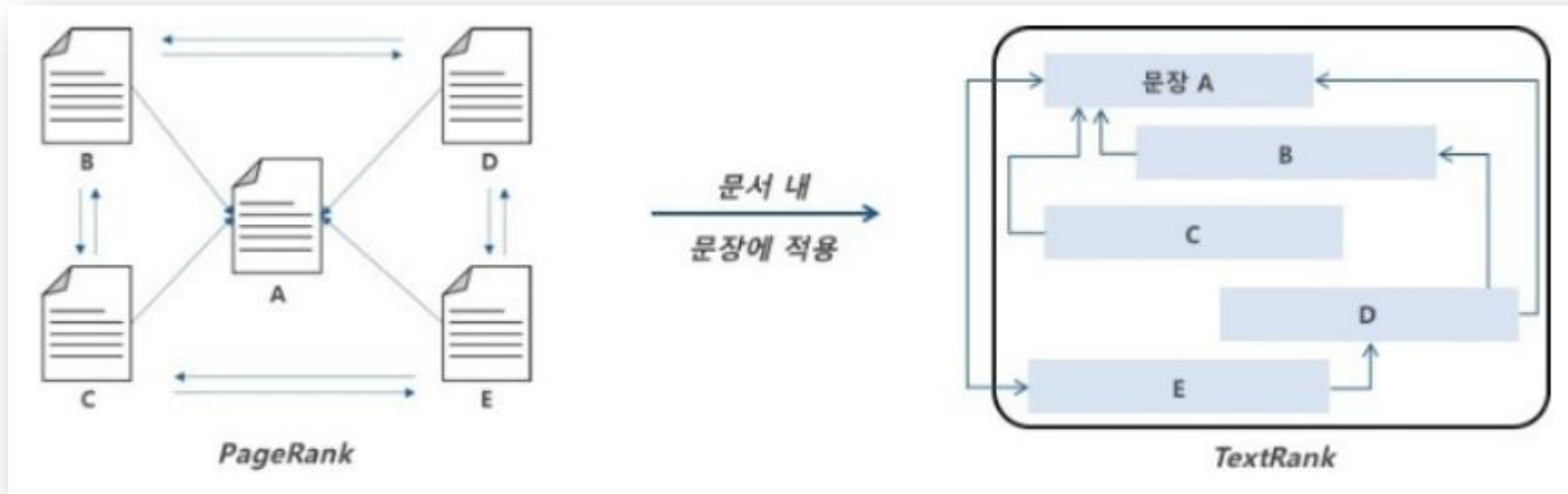
개발 현황

Text Rank란?

구글의 Page Rank를 Text에 적용시킨 알고리즘

Page Rank는 하이퍼링크를 가지는 웹 문서에 상대적 중요도에 따라 가중치를 부여하는 방법이다.

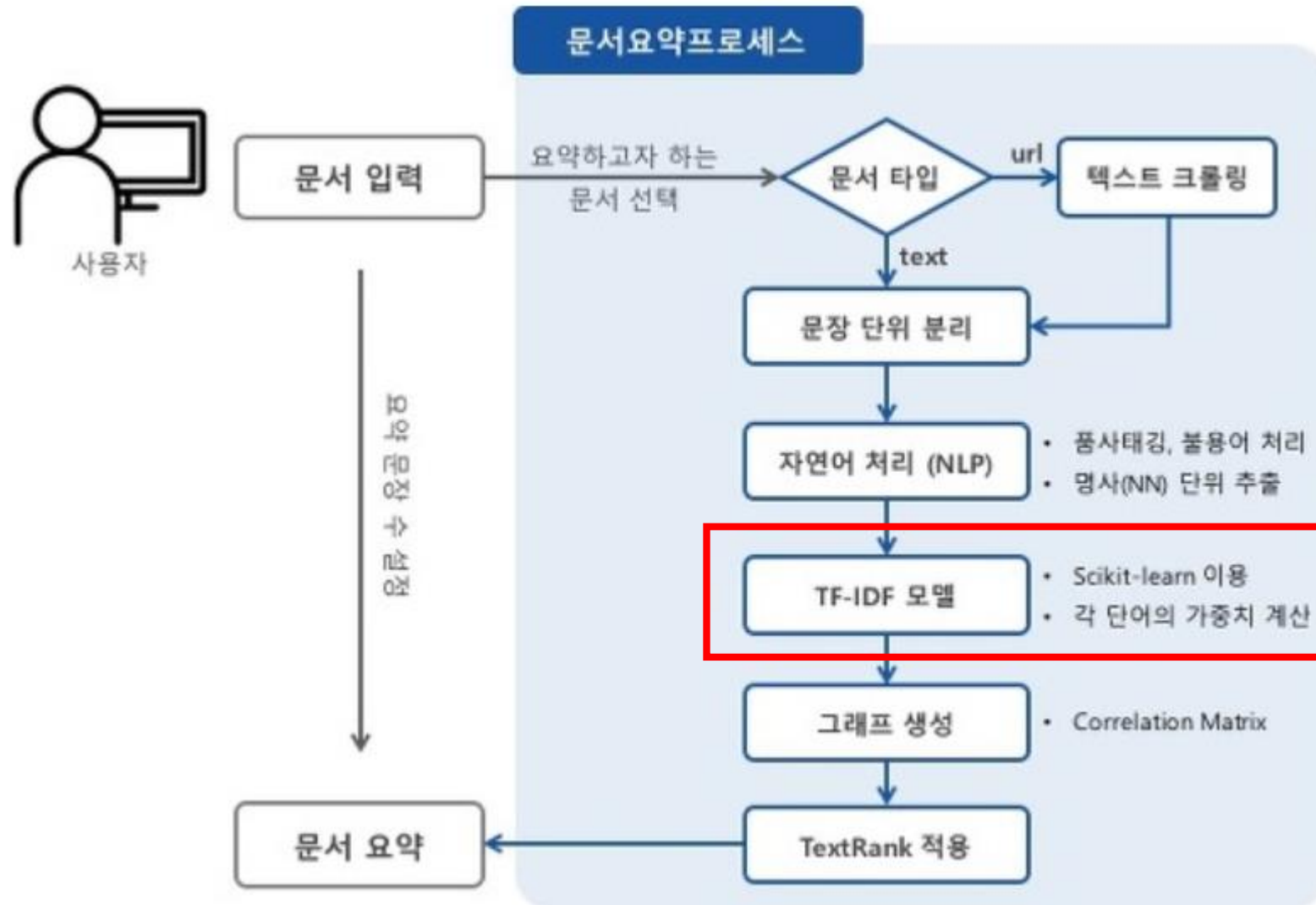
Text Rank는 Page Rank를 착안하여 문서 내의 문장을 이용하여 문장의 Ranking을 계산하는 알고리즘



개발 현황

- 기본 라이브러리를 참고하여 진행하면 프로그램 구현에 어려운 점은 없음.
- 하지만, 기본 라이브러리에서 keyword 추출 시 정치와 무관한 단어에 대해 가중치를 높게 잡히는 경향을 발견함.
- 결론 : 정치 관련 단어에 훨씬 높은 가중치를 부여하기 위해서 별도의 data를 수집하여 기본 라이브러리를 수정 중.

개발 현황



수정 및 개발 사항

개발 계획

심대범

- TFIDF
- Crawling
- 크롬 확장 프로그램

최승환

- Text Rank
- 메인 프로그램 개발
- 문서 및 보고서 작성

하태윤

- 감정 분석
- 메인 프로그램 개발
- 크롬 확장 프로그램

공동

정치 관련 DataSet 수집 (감정분석 DataSet, Text Rank DataSet)

개발 계획

개발내용	수행기간(주)													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
주제 선정							중 간 데 모	중 간 고 사						최 종 발 표
개발내용 구체화														
Dataset 수집														
KoNLPy 감정분석														
Text Rank 개발														
TFIDF 개발														
keyword 추출 비교 분석														
메인 프로그램 융합														
crawling														
크롬 확장프로그램 개발														
test 및 유지 보수														
추가기능 구현														
최종 보고서 작성														

| 질문 받습니다.

감사합니다.