

2022년 하반기

디지털/IT학회 프로그램

: 주식과 세상을 연결하다, STOCKNET

2조 | 고경주 박찬호 방미정 정찬빈 정혜나 최현수

삼성증권

CONTENTS

1 STOCKNET 개요

프로젝트 배경
STOCKNET 설명

2 STOCKNET 구현

노드 설정 : 종목별 언급량
엣지 설정 : 종목간 연결강도
프로토타입

3 기대효과 및 활용방안

기대효과
활용방안

4 발전 방향성

지속 가능한 방안 제시
개선점 및 보완점

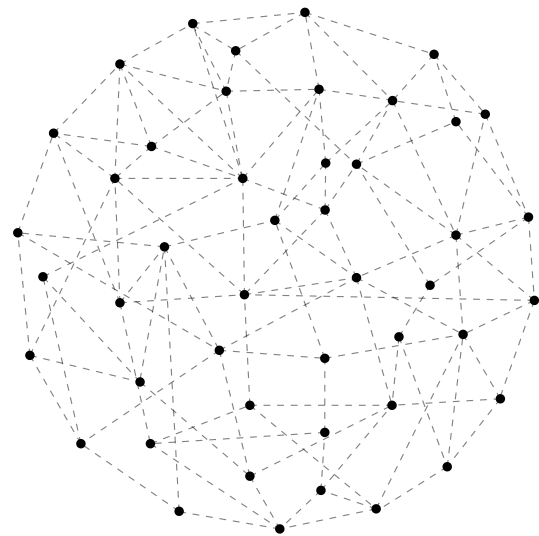
1

STOCKNET 개요

프로젝트 배경



주식 주가는 한 가지 요인으로 결정되는 것이 아니라 다양한 요인으로 결정됨
기업들 또한 사람들처럼 서로가 유기적으로 **관계**를 맺고 있고, 서로가 서로에게 영향을 받음



기업들 간의 **관계**를 **그래프(네트워크)** 형태로 시각화해서 보여줌으로써
투자자들에게 더 능동적인 투자 기회와 투자 **인사이트**를 제공하고자 함

STOCKNET 설명

기업 간 계약 관계에 민감하게 반응하는 주식 투자자들
기업들 간의 복잡한 상호작용을 통하여 끊임없이 변화하는 주식시장

➔ STOCKNET 종목 간의 연결관계를 한 눈에 볼 수 있도록 네트워크 형태로 시각화한 그래프
↳ KOSPI 200 기업 + 그 외 100개 상장기업

노드

상장기업 종목명

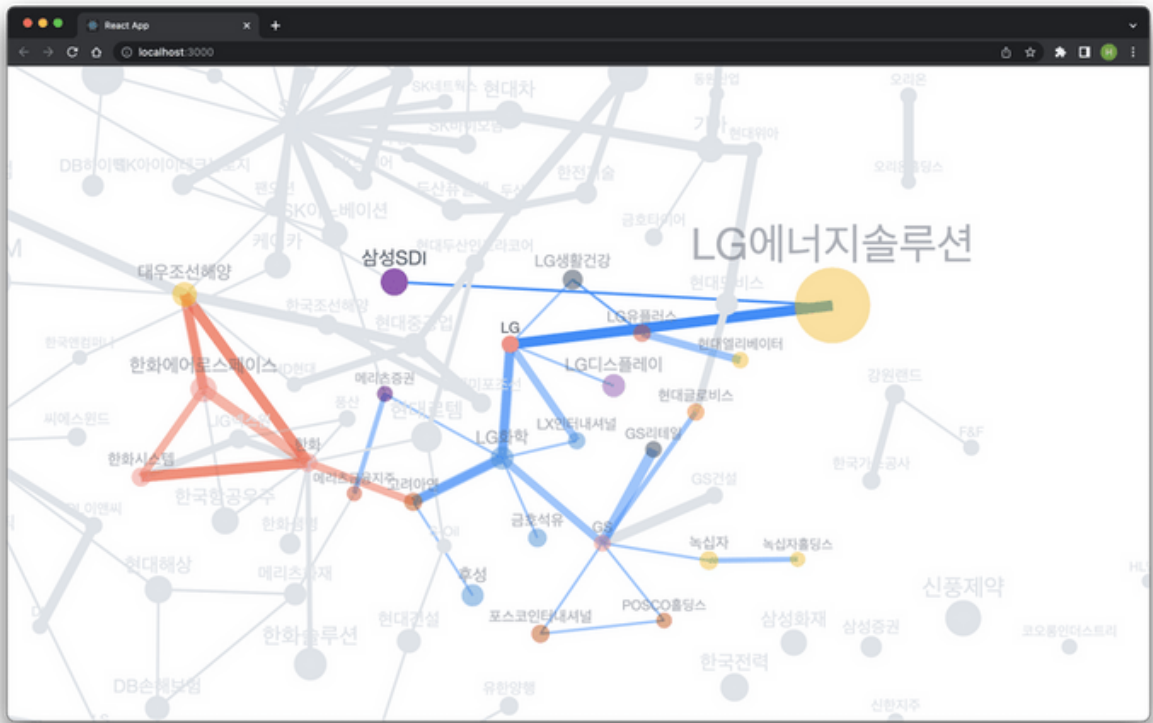
노드 크기: 네이버트렌드 검색량

노드 색깔: 섹터별로 구분

엣지

종목간 연결 강도

엣지 굵기: 계약 관련 뉴스에서의 두 기업의 동시 언급량



2

STOCKNET 구현

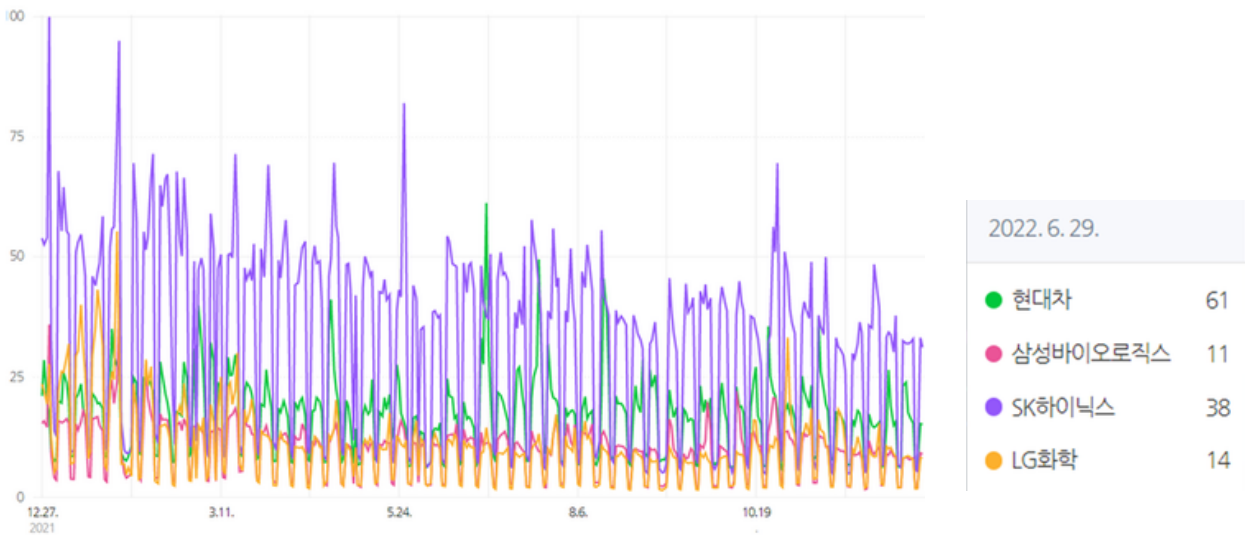
· **종목별 검색량**

[illegible]

활용 데이터

NAVER DataLab. 네이버 데이터랩 검색어 트렌드 데이터

검색어와 검색 횟수를 일간/주간/월간 단위로 추이를 확인할 수 있는 서비스
키워드 여러 개를 입력하면, 가장 많은 언급량(100)을 기준으로 상대적인 검색량 산출
* 키워드는 최대 5개까지 입력 가능



네이버 트렌드 데이터 API를 활용하여 일일 검색 데이터를 가져와서 사용

기간 2022.01.01-2022.12.26 키워드 KOSPI 200 기업 → 기업 5개씩 40묶음으로 검색

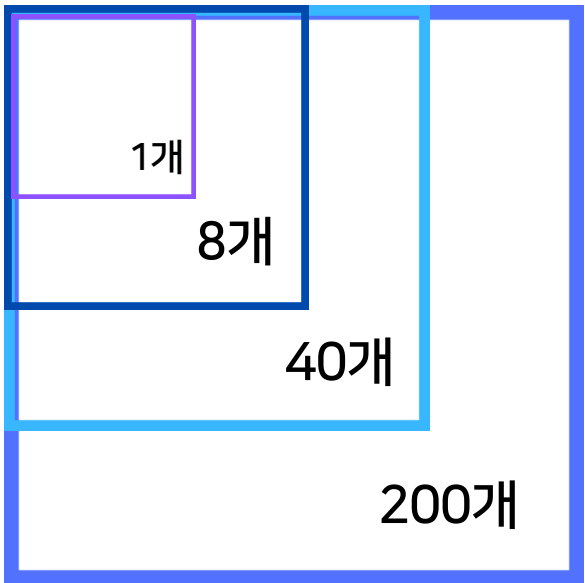
	삼성전자	삼성바이오로직스	SK하이닉스	LG화학	삼성SDI	현대차	NAVER	기아	POSCO홀딩스	셀트리온	카카오	삼성물산
0	11.26	0.60	2.51	0.77	1.15	2.44	12.64	3.11	0.00	2.82	3.73	0.43
1	11.31	0.53	2.34	0.87	1.45	2.63	12.67	3.49	0.00	6.28	4.37	0.44
2	59.31	2.42	12.14	3.33	6.40	6.60	22.67	5.58	0.00	3.54	3.28	0.37



언급량이 상대적이기 때문에 묶음끼리의 비교가 어렵다는 이슈 발생 → 뒷 장에서 문제 해결

구현 LOGIC

- 1 KOSPI 200 기업을 5개씩 40묶음으로 나누어 각 묶음의 1등 추출
- 2 40개 기업을 다시 8묶음으로 나누어 각 묶음의 1등 추출
- 3 8개 기업을 5개씩 뽑아 56개의 조합을 만들고, 트렌드지수가 가장 높은 1등 추출
- 4 1등을 기준으로 나머지 데이터들의 시계열 데이터를 하나의 데이터프레임으로 구조화



0	LG에너지솔루션	21	일진머티리얼즈
1	NAVER	22	CJ대한통운
2	카카오	23	신세계
3	신한지주	24	대우조선해양
4	하나금융지주	25	대한전선
5	카카오뱅크	26	DB하이텍
6	HMM	27	대우건설
7	대한항공	28	효성티앤씨
8	카카오페이	29	DL이앤씨
9	하이브	30	현대두산인프라코어
10	SK바이오사이언스	31	대웅
11	LG유플러스	32	신풍제약
12	한국항공우주	33	종근당
13	넷마블	34	아시아나항공
14	SK아이이테크놀로지	35	대상
15	삼성카드	36	KG스틸
16	한화에어로스페이스	37	명신산업
17	에스디바이오센서	38	케이카
18	팬오션	39	HDC현대산업개발
19	이마트		
20	현대해상		
21	일진머티리얼즈		

상위 40개 KOSPI기업

LG에너지솔루션	35
하이브	15
이마트	1
HDC현대산업개발	5
1등	LG에너지솔루션

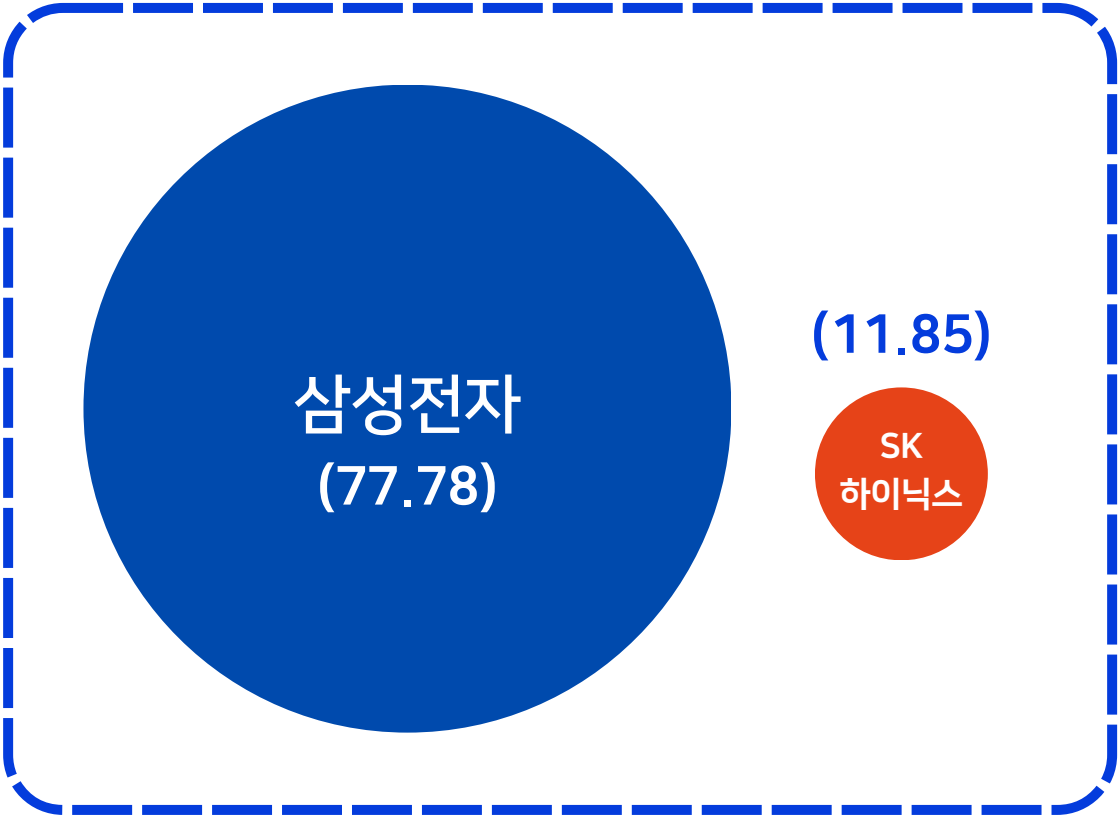
구현 LOGIC

5 종목별 검색량을 바탕으로 노드 크기 설정

	전체 합	전체 합 등수	평균
삼성전자	14039.50	1	77.78
삼성바이오로직스	536.80	109	2.97
SK하이닉스	2139.04	26	11.85
LG화학	554.30	103	3.07
삼성SDI	1814.05	34	10.05
현대차	1997.67	27	11.07
NAVER	7728.85	5	42.82
기아	1874.55	30	10.39
POSCO홀딩스	19.48	196	0.13
셀트리온	539.77	108	3.68
카카오	867.63	74	5.92
삼성물산	104.11	190	0.71
KB금융	1181.73	54	6.55
현대모비스	1276.50	51	7.07
신한지주	13337.40	2	73.89

	전체 합	전체 합 등수	평균
LG전자	2668.86	21	14.79
삼성생명	1289.06	50	7.14
SK이노베이션	1817.81	33	10.07
SK	860.64	75	4.77
한국전력	3446.70	14	19.10
포스코케미칼	2638.97	22	14.62
KT&G	839.07	79	4.65
하나금융지주	7734.88	4	42.85
LG	595.57	101	3.30
카카오뱅크	4249.26	9	23.54
고려아연	548.87	105	3.04
LG생활건강	850.24	76	4.71
SK텔레콤	1867.05	31	10.34
현대중공업	1450.31	44	8.03
HMM	3607.88	12	19.99

EXAMPLE



활용 데이터

NAVER NEWS 네이버 경제 뉴스 크롤링 데이터

기간 2022.01.01-2022.12.24 검색어 KOSPI 200 기업

	검색어	날짜	언론사	제목	요약	링크
0	메리츠 화재	3일 전	머니S	김용범 메리츠화재 대표, '영업통' 김경환 부사장 앉혔다... 이유는?	김용범 메리츠화재 대표가 김경환 전무를 부사장으로 승진시키는 등 지난 23일 임원인...	http://moneys.mt.co.kr/news/mwView.php?no=2022...
1	메리츠 화재	4일 전	매경이 코노미	내년 車 보험료 내린다...현대·KB·메리츠화재 등 줄 인하	롯데손보 2.9%로 인하폭 가장 커 메리츠 2.5%, 현대·KB 2.0% 등 현대해...	https://www.mk.co.kr/article/10578542

426,837 news



계약 관련 뉴스 필터링 → '요약' 컬럼에 계약 키워드가 들어간 뉴스만 필터링

계약 키워드 계약,체결,합병,인수,흡수,MOU,합작,협력,협약,협정,파트너쉽

426818	SK스 퀘어	2022.03.30.	아시아 투데이	SK하이닉스 인텔 키파운드리 이 어 ARM에 관심 사업 다각화 속도	박 부회장은 지난 28일 SK스퀘어 주주총회에 서도 반도체 업체는 규모가 큰 곳부터...
426822	SK스 퀘어	2022.12.14.	워크투 데이	하나은행 버추얼 휴먼 솔루션 기업 온마인드와 전략적 업무협약 체결	황보현우 하나은행 데이터 제휴투자본부 본부 장은 온마인드와 업무제휴를 통해 급변하는...

48,653 news

한국경제 PICK | A12면 1단 | 1일 전 | 네이버뉴스
[단독] 한화, 대우조선해양 이어 STX중공업도 인수 추진

구현 LOGIC

- 1
- CKONLPY의 OKT 형태소 분석기를 이용해 '요약' 컬럼 토큰화
- ↳ 한 단어로 인식할 단어를 미리 지정 가능 → 2,492개의 상장기업명 입력

- 2
- 뉴스에 등장하는 상장기업명 추출

	검색어	날짜	언론사	제목	요약	링크	token	pos	noun	company
37996	현대글로비스	3일 전	동아일보	단독 삼아알미늄 LG에너지솔루션과 양극박 장기 공급 추진 8년간 23조 규모	현대자동차그룹이 지난 2018년 지배구조 개편을 시도하면서 현대모비스와 현대글로비스...	NaN	[현대자동차, 그룹, 이, 지난, 2018년, 지배구조, 개편, 을, 시도, 하면서...]	[(현대자동차, Noun), (그룹, Noun), (이, Jose), (지난, No...]	[현대자동차, 그룹, 지난, 지배구조, 개편, 시도, 현대모비스, 현대글로비스, 분...]	[현대자동차, 현대모비스, 현대글로비스, 삼아알미늄]



- 3
- KOSPI 200 기업별로 언급되는
상장기업명 개수 카운트

	word	count		word	count
0	현대모비스	89	0	현대제철	141
1	GS	46	1	SG	110
2	현대자동차	33	2	OCI	61
3	쏘카	33	3	LG에너지솔루션	43
4	레이	26	4	레이	42
5	SK	24	5	롯데케미칼	37
6	한화솔루션	21	6	삼성엔지니어링	34
7	삼성물산	20	7	ISC	32

...

현대글로비스

포스코케미칼

삼성증권

구현 LOGIC

4 STOCKNET 노드에 추가할 100개 기업 선정

→ KOSPI 200에 속하지 않는 2,292개 상장기업 중 언급량이 많은 상위 100개

최종 노드: 기존의 KOSPI 200 + 추가된 100개 기업

word count			word count		
0	SG	1453	16	롯데칠성음료	214
1	레이	1313	17	KG케미칼	211
2	현대자동차	456	18	CS	209
3	태양	352	19	쌍방울	189
4	국동	351	20	선진	187
5	DB	337	21	현대리바트	185
6	금호석유화학	336	22	에스엘바이오닉스	182
7	쌍용자동차	304	23	STX중공업	180
8	NEW	301	24	디오	179
9	한국전력공사	291	25	E1	177
10	KD	275	26	카페24	175
11	코오롱인더	267	27	KG이니시스	175
12	연우	243	28	테스	158
13	성일하이텍	243	29	코오롱	157
14	영풍제지	226			

추가된 100개 기업

5 종목간 상호언급량을 기준으로 엣지 굵기 설정

→ A기업 뉴스에서의 B기업 언급량 + B기업 뉴스에서의 A기업 언급량

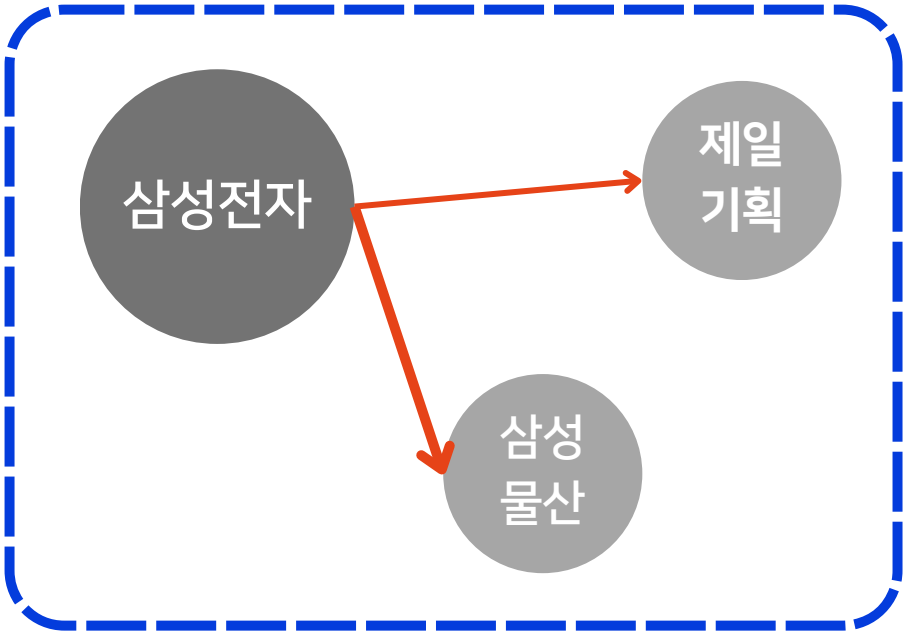
word count			word count		
0	현대제철	141	0	포스코케미칼	113
1	SG	110	1	한국철강	11
2	OCI	61	2	동국제강	8
3	LG에너지솔루션	43	3	대성	8
4	레이	42	4	현대자동차	8
5	롯데케미칼	37	5	SG	4
6	삼성엔지니어링	34	6	DB	4
7	ISC	32	7	에코프로에이치엔	3

포스코케미칼

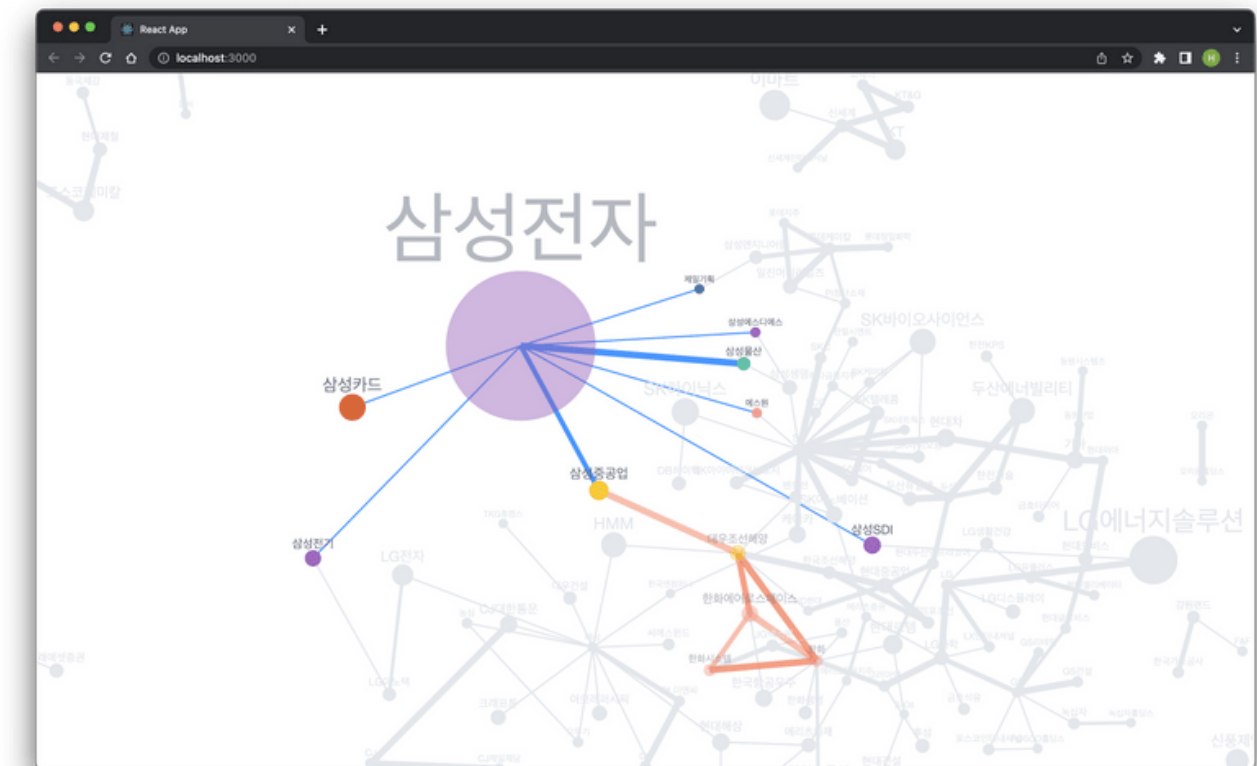
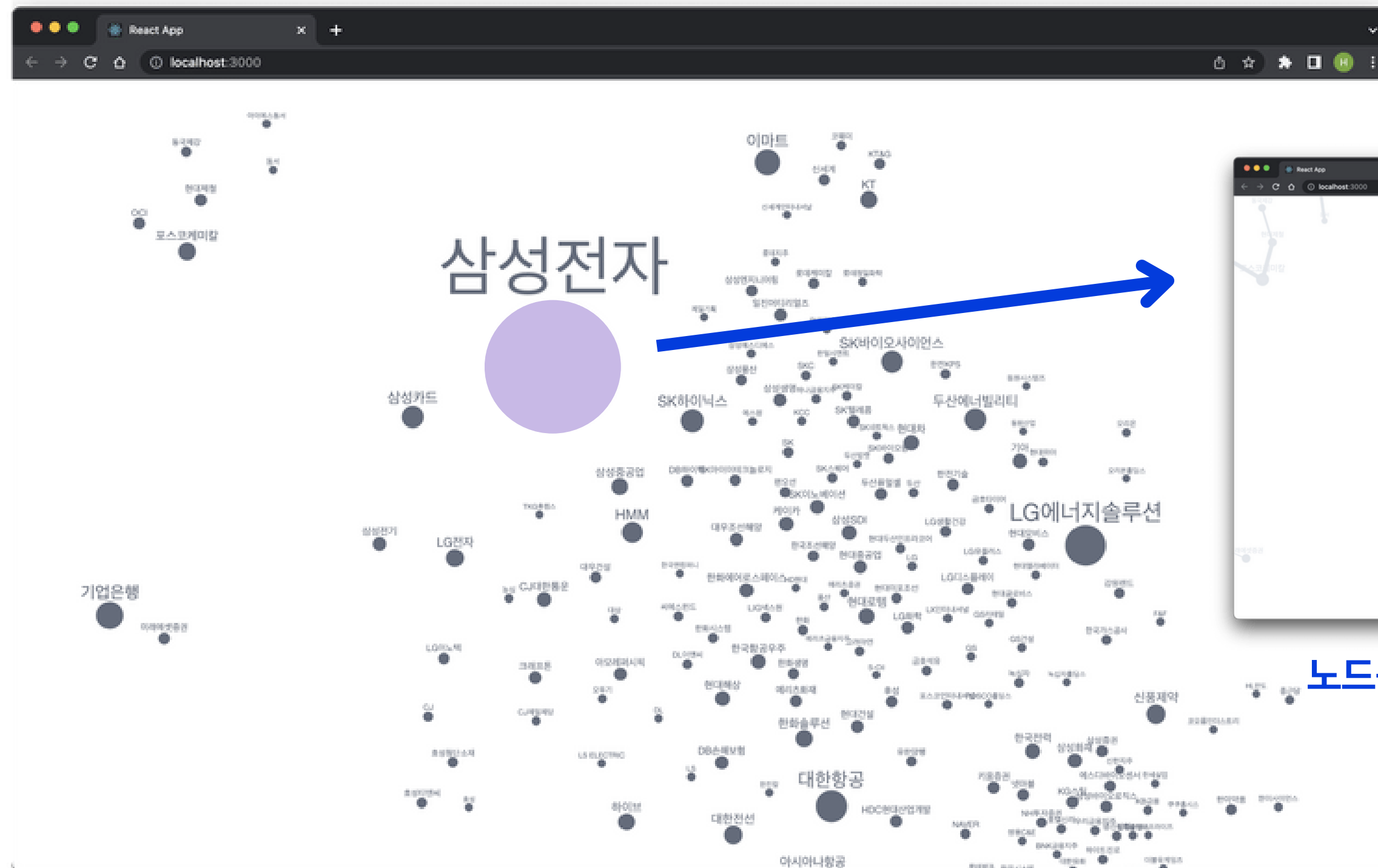
현대제철

→ 포스코케미칼과 현대제철의
연결 강도 : 141+113 = 254

EXAMPLE



STOCKNET



노드를 클릭하면 연결된 관계들이 보임

삼성전자



프로토타입

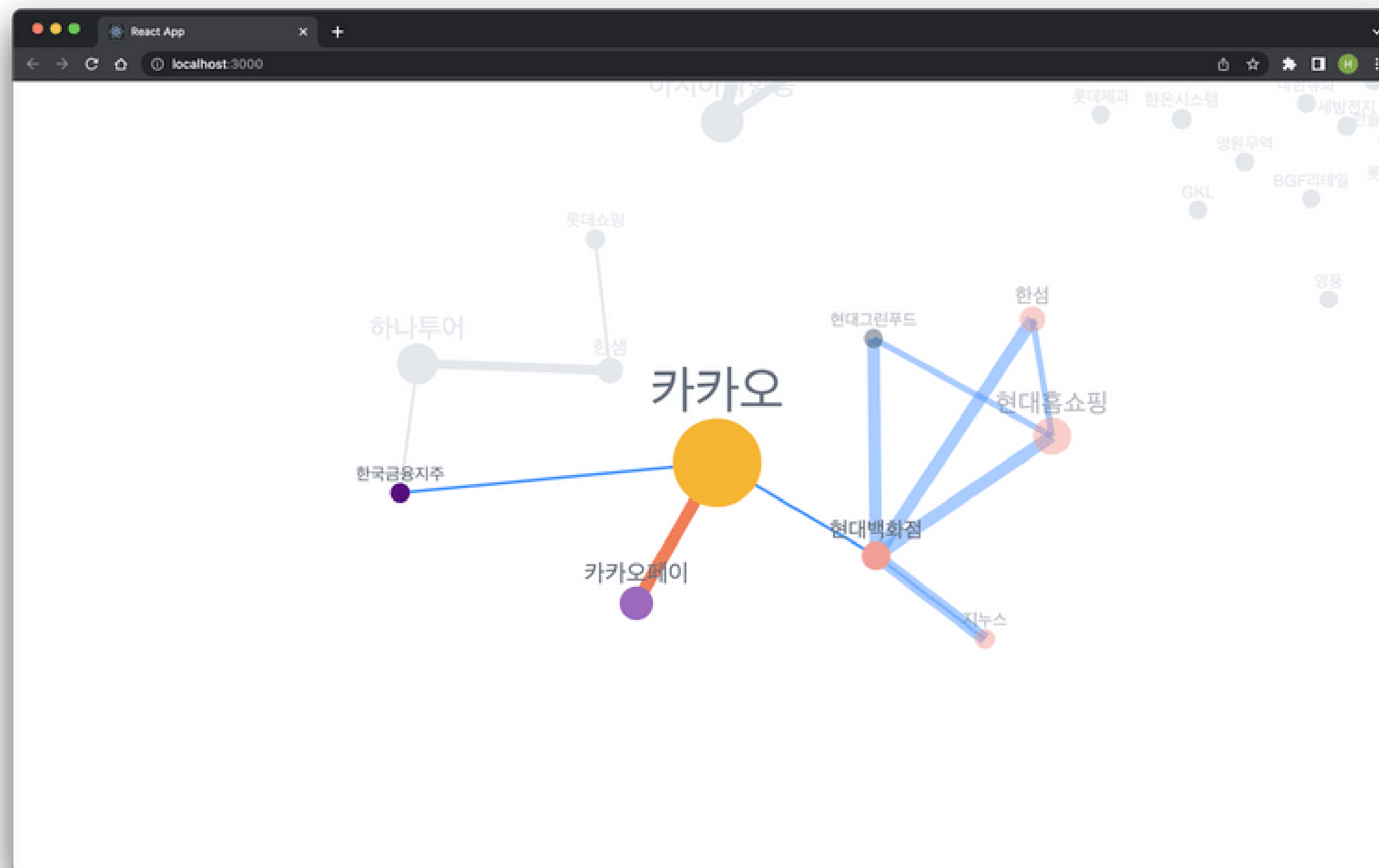
카카오

아시아경제 | 5일 전 | 네이버뉴스

한국투자증권, 3.4조 투자해 카카오뱅크 지분 취득

브레이크뉴스 | 2022.07.05.

현대백화점면세점, 카카오엔터테인먼트와 'K-컬처 콘텐츠 제휴 협약' 체결



삼성증권

프로토타입

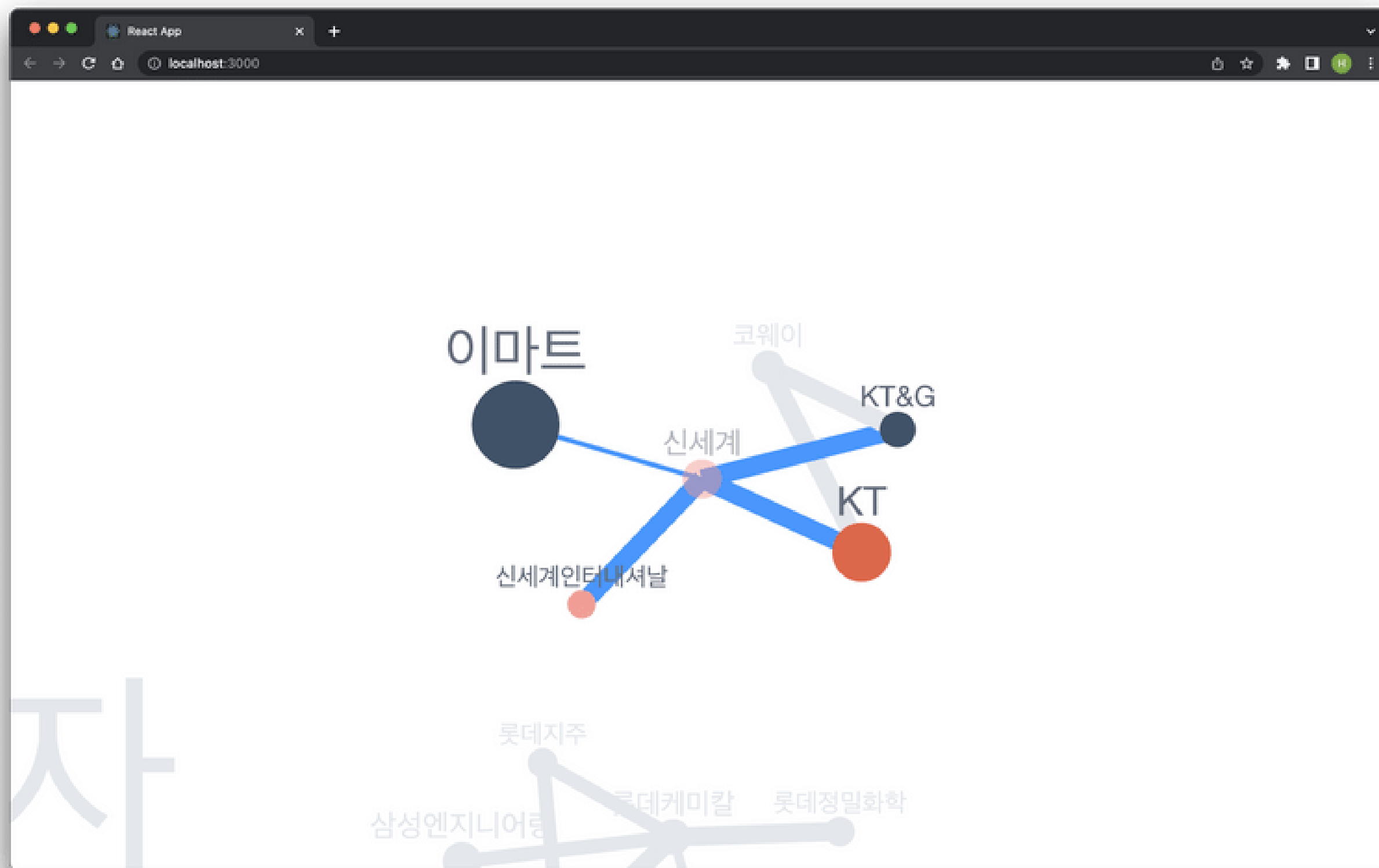
이마트

디지털데일리 | 2022.12.14. | 네이버뉴스

KT-신세계, 디지털 생태계 조성 '맞손'...사업협력체 조직

뉴스1 | 2022.12.14. | 네이버뉴스

신세계그룹 멤버십으로 **KT** 요금 할인 받는다...사업 협약 체결

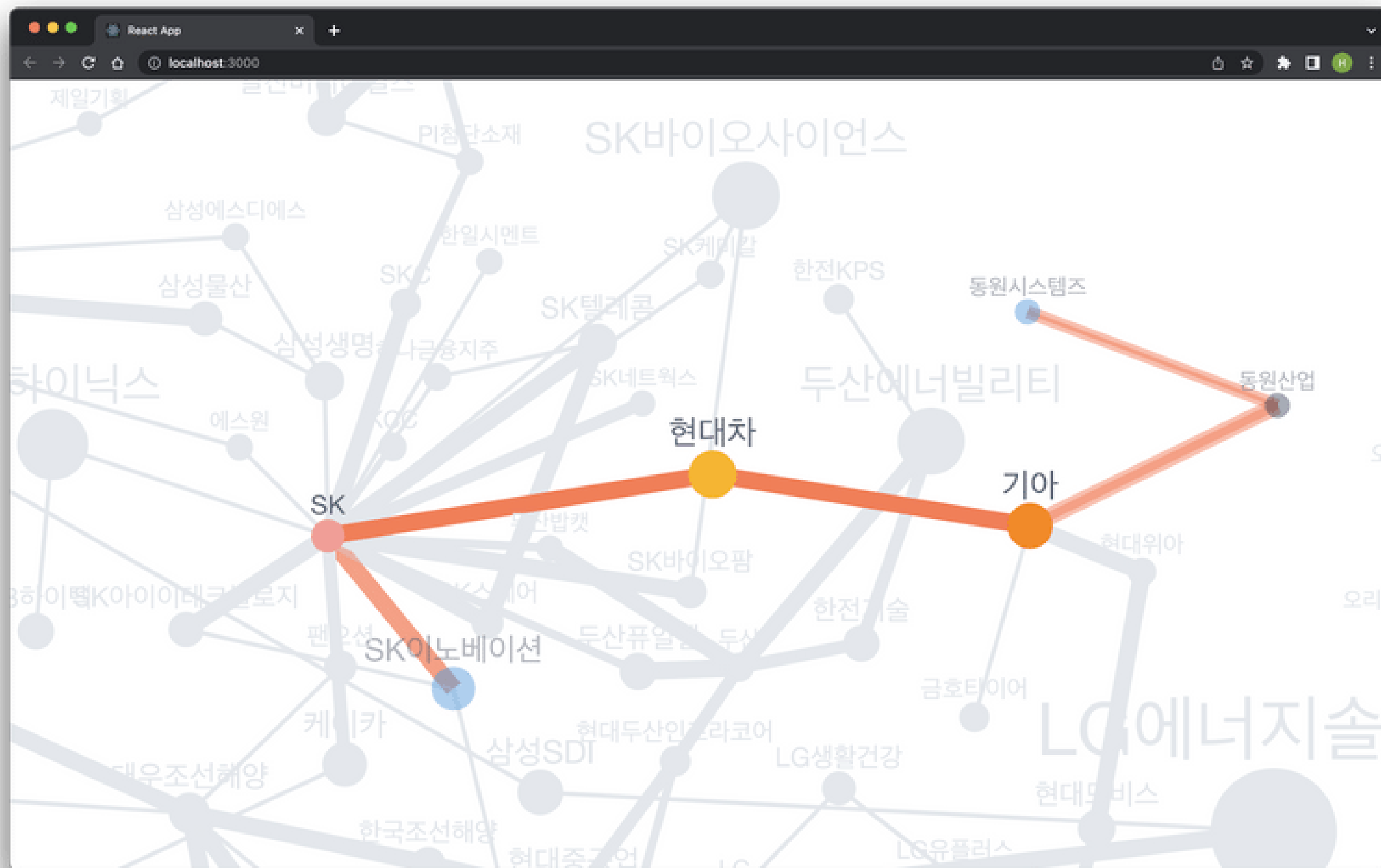


삼성증권

현대차

현대·SK온, 美 조지아주 애틀랜타에 배터리 합작공장 발표

현대차그룹-SK온 맞손... 북미 배터리 공급 MOU 체결



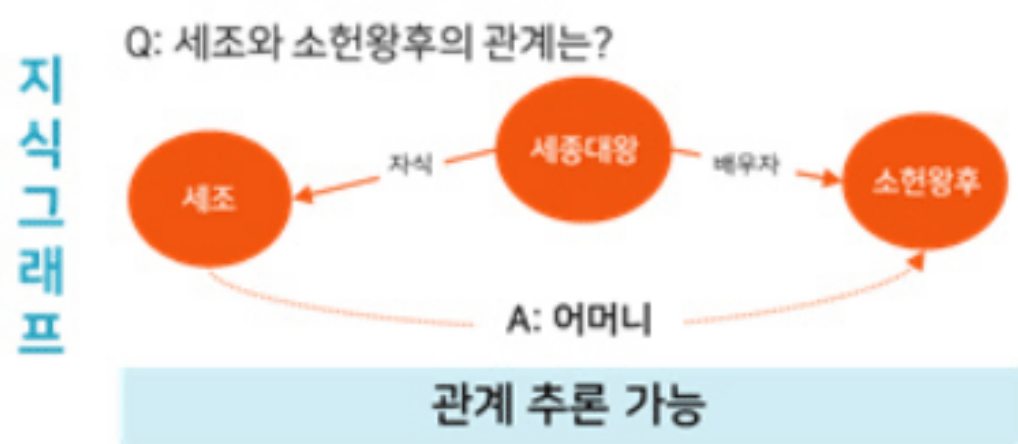
삼성증권

3

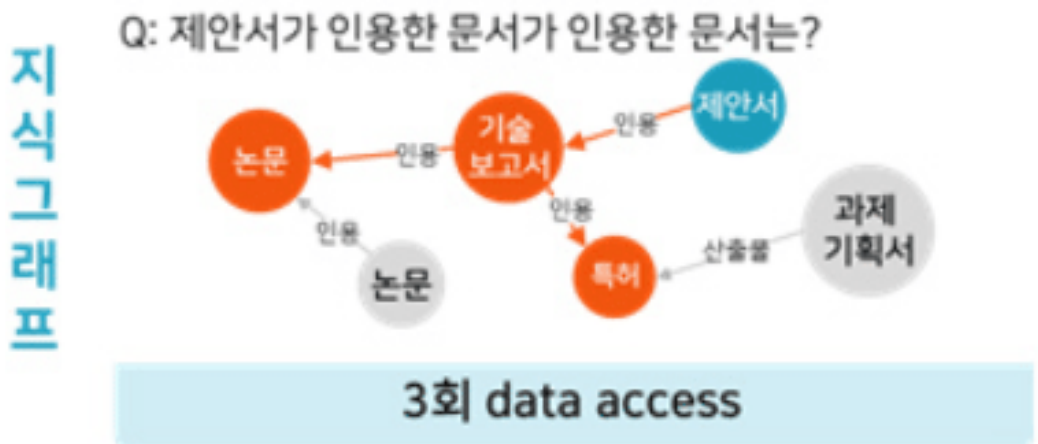
활용방안 및 기대효과

기대효과

Reasoning



Performance



Flexibility



기존 서비스

관계 테이블

인물 A	인물 B	관계
세종대왕	소헌왕후	배우자
세종대왕	세조	자식

정의되지 않은 관계 유추 불가능

기존 서비스

문서 테이블		인용 테이블	
ID	Doc	DocID	RefID
1	제안서	1	2
2	기술 보고서	2	3
3	특허	2	5
4	논문	4	6
5	논문	5	4

12회 data access

기존 서비스

문서 테이블		인용 테이블		심사 (새 테이블)	
ID	Doc	DocID	RefID	PER ID	RefID
1	제안서			6	3
2	기술 보고서	1	2		
3	특허				
5	논문	2	3		
				작성 (새 테이블)	
ID	이름	PER ID	RefID		
6	홍길동	5	6	6	5

데이터 변경에 복잡한 과정 필요

- 1 특정 종목들 간의 유사성 및 언급량에 대한 정보를 쉽게 보여주며, 데이터의 수정 및 관리에 훨씬 용이함
- 2 한눈에 들어오는 데이터들의 복합적 배열로 투자자로 하여금 정보 탐색 과정에서 원활한 추론 및 시간의 단축을 이끌어 냄

활용방안

엣지 활용

KOSPI + KOSDAQ에 상장된
약 2,600여 개 기업들의
관계도 파악 가능

- 1) 기업 간의 포괄적 협력 관계 파악 가능
ex) 원청 - 하청 관계
- 2) 자회사 - 모회사 구조의 표시
: 지배 구조 파악에 용이할 것
ex) LG화학 - **LG에너지솔루션** - 고려아연

노드 활용

종목별 색 지정을 통한
노드별 섹터 구분 가능

- 1) 같은 색 노드는 같은 섹터를 의미
: 종목별 산업군 파악 및 동종 업계 분석에 용이
- 2) 서로 다른 섹터들 간의 연결을 바탕으로
산업군 간의 연관성 파악
ex) 코로나19 - 물류 업계 - **골판지 회사**

시계열 활용

연도 내 분기별 STOCKNET을
작성하여 산업의 주기성 파악

- 1) 특정 종목 및 섹터의 분기별 영향력 비교
: 해당 산업군의 주기성 파악
- 2) 서로 다른 종목 간의 연결성 변화 파악
ex) **에너지** 업계 - **가전 제품** 업계

4

발전 방향성

지속 가능한 방안 제시

WORKFLOW 자동화



KOSPI200 리스트 수집

KRX 정보데이터시스템 제공
[12005] 전종목 기본정보 수집



뉴스 크롤링 & 분석

KRX 종목명 기반 뉴스 크롤링 후
NLP로 토큰화하여 언급량 분석



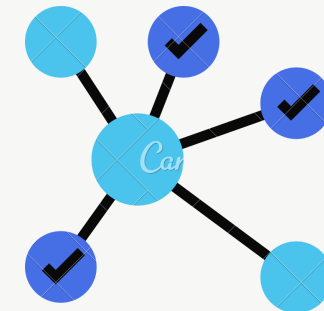
트렌드 상대지수 도출

네이버 트렌드 API 활용하여
KOSPI200 트렌드 상대지수 산출



DB 저장

DB에 company, code,
trend, sector 실시간 업데이트



언급&검색량 시각화

종목 별 섹터정보, 검색량, 종목 간
동시 언급량 지식그래프로 시각화

프로젝트 지속가능성

실시간 데이터 수집(종목, 뉴스, 검색어 트렌드) -> 기업명 사전을 통한 언급량 분석 -> DB 업데이트 -> 언급&검색량 지식그래프 시각화까지의 Workflow를 자동화하여, 데이터 수집과 분석 코드의 수정을 위한 인력투입 및 비용의 발생 없이 실시간 데이터 업데이트 및 프로젝트 규모 확장 가능

개선점 및 보완점

트렌드 지수를 활용 할 때, SNS랑 유튜브 등은 어떻게 활용할 수 있을 것인가

해외 시장(기업)의 영향력 또한 반영해야 할 것

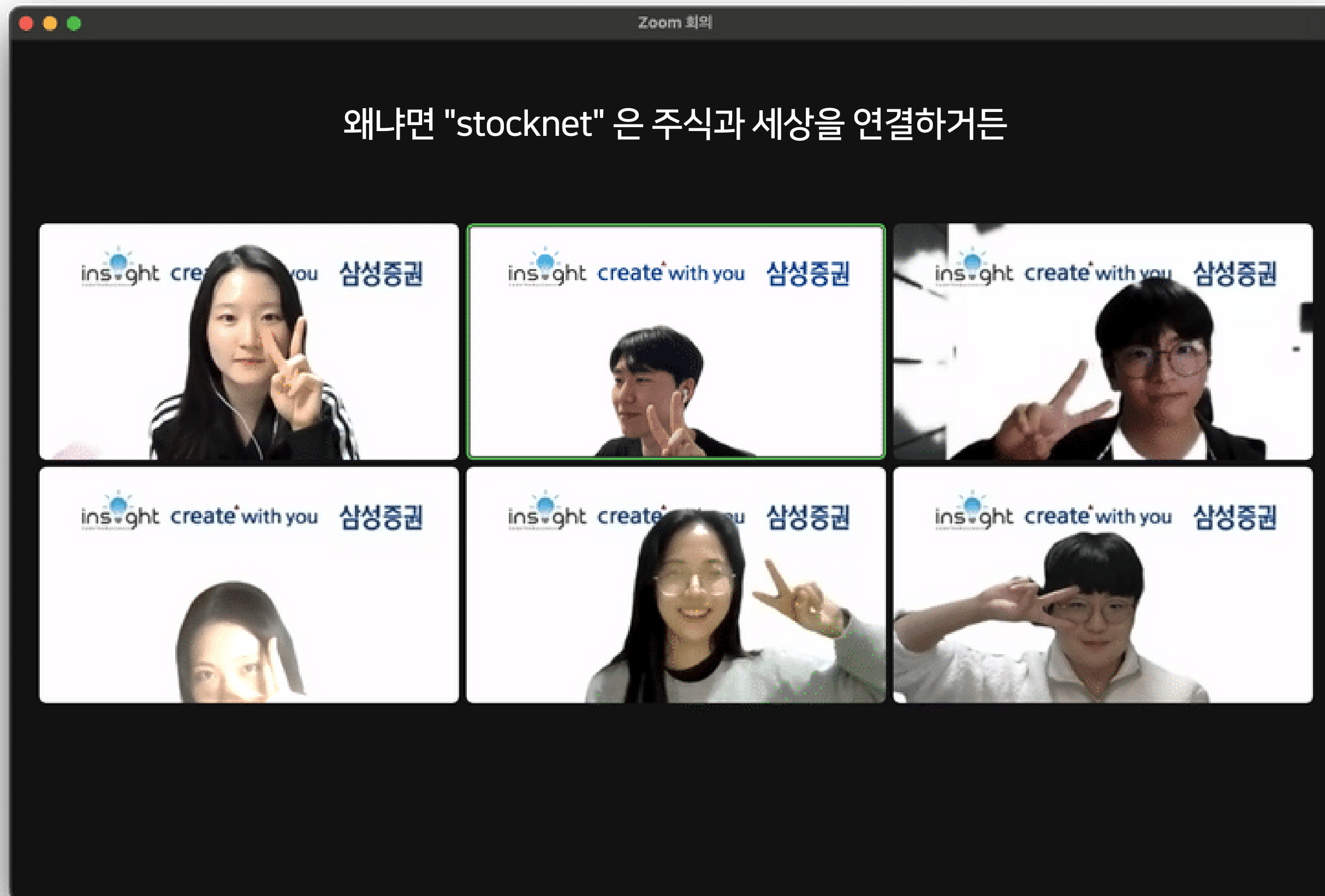
'썸트렌드', '구글 트렌드' 등의 다양한 지표 또한 활용할 수 있을 것

뉴스 내에서 계약, MOU, 인수 등의 언급 뿐만 아니라 규모 또한 반영한다면 더 정확할 것

비상장 기업들의 관계도는 어떻게 파악할 수 있을 것인가

투자자들이 STOCKNET의 엽지를 클릭 시, 관련 뉴스 요약 및 링크를 열람할 수 있게 하는 것은 어떠한가

2조





감사합니다

APPENDIX 00

개발환경

제출 코드 파일명
OS, 라이브러리 버전

데이터 수집 & 분석

file : 뉴스크롤러_000_to_199.ipynb

- 개발환경 windows10
- Pandas Version : 1.3.4
- Requests Version : 2.26.0
- BeautifulSoup : 4.10.0

file : Data_Analysis.ipynb

- 개발환경 windows10
- Pandas Version : 1.4.2
- Numpy Version : 1.21.5
- Urllib.request Version: 3.9
- json Version : 2.0.9
- openpyxl Version : 3.0.9
- matplotlib Version : 3.5.1

folder : stocknet_edge

- 개발환경 windows10
- Pandas Version : 1.4.2
- Numpy Version : 1.21.5

프로토타입

folder : stocknet -prototype

- 개발환경 : macOS Ventura
- node.js v16.15.1
- React v18.2.0
- Notion API
- Cytoscape.js

APPENDIX 01

데이터 수집

- KRX 전체종목정보
- 네이버 트렌드 지수
- 2022 네이버 뉴스

데이터 수집

KRX 정보데이터시스템

KRX 정보데이터시스템 - selenium을 활용하여 KOSPI200 종목정보 조회 및 다운로드 과정 자동화

네이버데이터랩 - 네이버 통합검색어 트렌드 API

파라미터 : startDate, endDate, timeUnit, keywordGroups, keywordGroups.groupName, keywordGroups.keywords, gender, ages

응답값: startDate, endDate, timeUnit, results.title, results.keywords, results.data.period, results.data.ratio

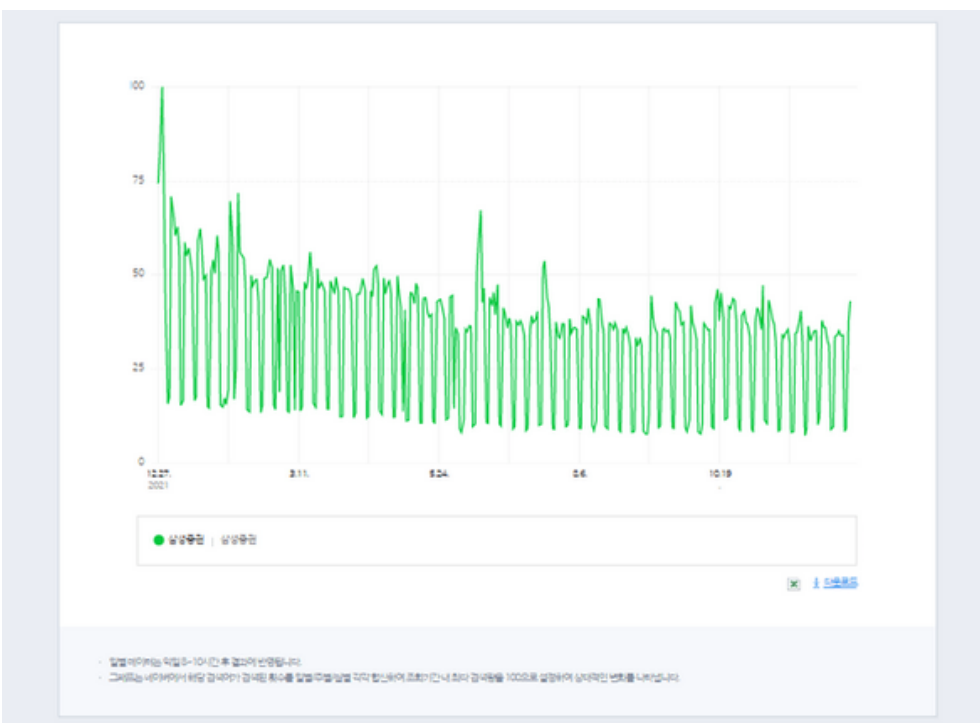
```
import os
import sys
import urllib.request
client_id = "YOUR_CLIENT_ID"
client_secret = "YOUR_CLIENT_SECRET"
url = "https://openapi.naver.com/v1/datalab/search";
body = "{W\"startDateW\":W\"2017-01-01W\",W\"endDateW\":W\"2017-04-30W\",W\"timeUnitW\":W\"monthW\",W\"keywordGroupsW\":[{W\"groupNameW\":W\"한글W\",W\"keywordsW\":[W\"한글W\",W\"koreanW\"]},{W\"groupNameW\":W\"영어W\",W\"keywordsW\":[W\"영어W\",W\"englishW\"]}],W\"deviceW\":W\"pcW\",W\"agesW\":[W\"1W\",W\"2W\"],W\"genderW\":W\"fW\"}";
```

네이버뉴스 크롤링 - 키워드기반 URL 접근, BeautifulSoup와 requests로 웹 파싱

시작날짜, 마지막날짜, 키워드, 페이지번호를 URL에 넣어 네이버뉴스 검색 페이지에 접근, bs4와 requests 활용해 HTML 소스 분석

```
for number in num_list:
    URL = 'https://search.naver.com/search.naver?where=news&sm=tab_pge&query={search_word}&sort=0&photo=0&field=0&pd=3&ds={startdate}&de={enddate}&cluster_rank=37&mynews=0&office_type=0&office_section_code=0&news_office_checked=&nso=so:r,p:from{start_date}to{end_date},a:all&start={page_num}'
    headers = {'user-agent':'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/88.0.4324.150 Safari/537.36',
               'accept':'text/html,application/xhtml+xml,application/xml;q=0.9,image/avif,image/webp,image/apng,*/*;q=0.8,application/signed-exchange;v=b3;q=0.9'}
    url = URL.format(search_word = my_keyword ,startdate= my_start_date1,enddate = my_end_date1,
                    start_date = my_start_date2,end_date = my_end_date2 ,page_num = number)
```

삼성증권



N 네이버 데이터랩

통합 VIEW 이미지 지식IN 인물루어서 동영상 쇼핑 뉴스 어학사전 지도

관한도순 최신순 오래된순

PICK 언론사가 선정한 주요가사 혹은 심층기획 기사입니다.

올해형 6시간 전

[스타트업 단신] '에이블리코퍼레이션, 마이프랜차이즈, 카닥, 휘...'

업 사용자 빅데이터 기반 '2022 에이블리 연말 결산' 에이블리코퍼레이션이 운영하는 스타일 커머스... 컴퍼니케이프트네스, 네이버 D2SF, 라구나인베스트먼트 등...

국방일보 5일 전

시데이터 기반 국방혁신 4.0 속도 낸다

하정우 디지털콘텐츠정부위원회 시데이터본고장(데이터에 몰로바 AI업 연구소장)은 "다양한 국방 데이터를 전략적으로 활용하기 위해서는 민간 전문역량 활용이 절..."

국방일보 5일 전

시데이터 기반 국방혁신 4.0 속도 낸다

하정우 디지털콘텐츠정부위원회 시데이터본고장(데이터에 몰로바 AI업 연구소장)은 "다양한 국방 데이터를 전략적으로 활용하기 위해서는 민간 전문역량 활용이 절..."

연관 검색어

네이버데이터랩 네이버 데이터센터

네이버 키워드 도구 네이버 검색광고 키워드

네이버 검색어 트렌드 네이버 분석

코로나19 현황

확진 현황 및 백신 접종 정보 안내

신속항원검사

내 주변에서 신속하게 검사 받기

당장 떠나고 싶다면 국내 여행부터!

지금 가장 핫한 국내 여행지로 떠나보세요