

감염병 전파 예측 시스템

졸업프로젝트

고길재 | 김민혁 | 조수빈

01

02

Contents

03

04

05

06

07

08

09

10

1. 프로젝트 배경

2. 프로젝트 목표

- Success Criteria

3. 프로젝트 요구사항

- 기능 요구사항, 품질 요구사항

- System test case

- Traceability Matrix

4. 프로젝트 설계

- Component diagram

- Sequence diagram

5. 1st Iteration

- System testing Result

6. 2nd Iteration

- System testing Result

- Pass/Fail Criteria

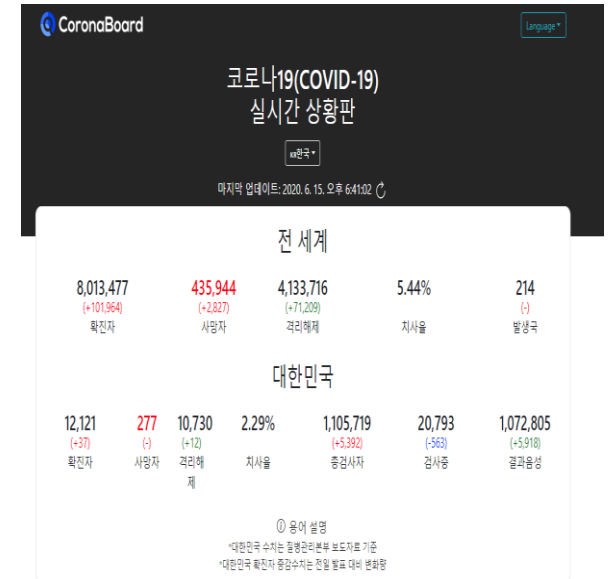
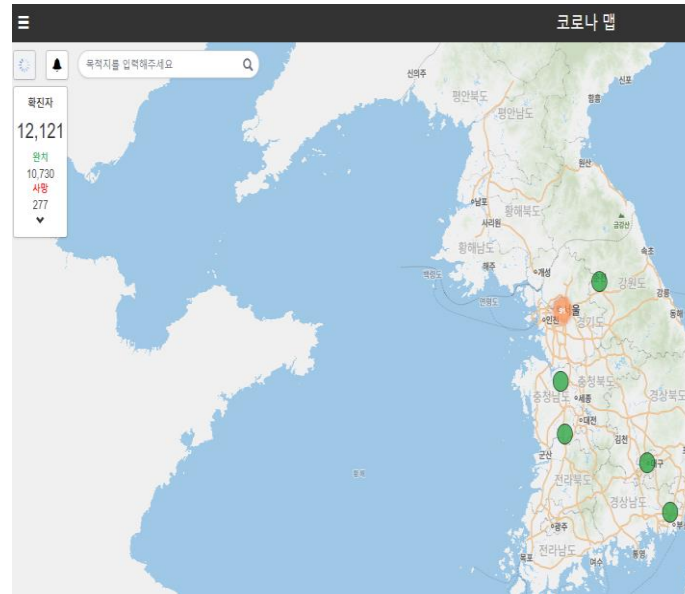
7. 3rd Iteration

- System testing Result

- Pass/Fail Criteria

프로젝트 배경

Covid19 현황을 보여주는 다양한 서비스 존재하지만
언제 종식될 지 알려주는 서비스는 없음



01
02
03
04
05
06
07
08
09
10

프로젝트 목표

Description

확진자 자료를 이용해 감염병 전파를 예측하는 웹 애플리케이션

전 세계 Covid-19 감염 데이터를 크롤러를 이용해 자동으로 수집한 뒤 국가별 유행률을 계산하여 Covid-19 확진자 추이 예측 및 종료시기 제공

기능

- 국가별 확진자 현황 및 전일 대비 변화량 제공
- 감염병 발생일부터 현재까지 확진자 그래프 제공
- SIR / SEIR 모델과 데이터 피팅 알고리즘을 이용한 확진자 예측
- 국가별 Covid-19 예상 종료시기 및 확진자 추이 예측 그래프 제공

목표

새로운 전염병 발생시 빠르게 적용 가능한 시스템 구축

01

Success Criteria

02

03

04

구현한 SIR / SEIR 모델이 얼마나 유용한지 판별하기 위해
결정계수(R^2) 0.5 이상을 목표

05

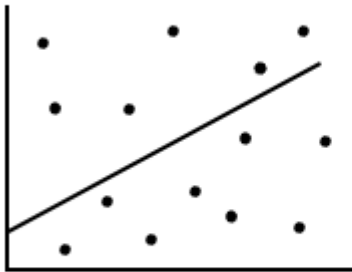
06

07

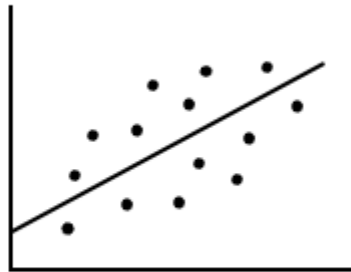
08

09

10



$$R^2 = 0$$



$$R^2 = 0.5$$



$$R^2 = 1$$

01
02
03
04
05
06
07
08
09
10

프로젝트 요구사항

기능 요구사항

1. 데이터 수집

- 1.1. 약 210개국의 국가별 확진자 수, 사망자 수, 회복 인원을 크롤링 한다.
- 1.2. 크롤링한 데이터를 시계열 형태로 가공한다.
- 1.3. 가공된 데이터를 데이터 베이스에 저장한다.
- 1.4. 24시간에 한 번씩 실행된다.

2. 데이터 가공

- 2.1. 데이터 베이스에 저장된 데이터를 불러와 모델의 Input에 알맞은 형태로 가공한다.
(날짜, 전체 인구, 확진자 수, 사망자 수, 회복 인원) -> (날짜, 감염 가능군, 감염군, 회복군)
- 2.2. 데이터 피팅을 통해 전파율, 회복율을 계산한다.
- 2.3. 전파율, 회복율을 이용해 유행 종료일까지의 확진자 수를 예측한다.
- 2.4. 성능 평가를 위한 결정 계수를 계산한다.

3. 데이터 시각화

- 3.1. 원하는 국가를 검색하여 전염병 데이터를 확인할 수 있다.
- 3.2. 국가별 전염병 확진자 데이터를 표의 형식으로 보여준다.
- 3.3. 국가별 전염병 확진자 추이를 그래프로 보여준다.
- 3.4. 국가별 전염병 종식 일자를 제공한다.

품질 요구사항

1. 사용자 편의성을 위해 UI가 직관적이어야 한다.
2. 새로운 전염병에도 사용할 수 있는 모델이어야 한다.

System Test Case

#	Description
Test 1	국가별 확진자 수, 사망자 수, 회복 인원을 성공적으로 크롤링 하는지 테스트한다.
Test 2	크롤링한 데이터를 시계열 형태로 가공하는지 테스트한다.
Test 3	크롤링한 데이터를 데이터 베이스에 성공적으로 저장하는지 테스트한다.
Test 4	24시간에 한 번씩 실행되는지 테스트한다.
Test 5	데이터 베이스에서 불러온 데이터가 모델에 알맞은 형태로 가공되는지 테스트한다.
Test 6	ODE(Ordinary Differential Equation) Solver의 결과값이 계산되는지 테스트한다.
Test 7	Matrix Class를 이용 행렬의 기본적인 연산(+, -, *, inverse, transpose, 행렬식)이 되는지 확인한다.
Test 8	데이터를 ODE Solver를 이용해 피팅 후 전파율과 회복율이 계산되는지 테스트한다.
Test 9	계산된 전파율과 회복율을 ODE Solver에 적용하여 전염병 유행 종료시기를 예측할 수 있는지 테스트한다.
Test 10	계산된 결정 계수가 Success Criteria를 만족하는지 테스트한다.
Test 11	국가를 검색했을 때 해당 국가의 전염병 데이터가 웹에 출력되는지 테스트한다.
Test 12	국가를 선택했을 때 해당 국가의 전염병 확진자 데이터가 표의 형식으로 웹에 출력되는지 테스트한다.
Test 13	국가를 선택했을 때 해당 국가의 전염병 확진자 추이가 그래프 형식으로 웹에 출력되는지 테스트한다.
Test 14	국가를 선택했을 때 지도 위에 전염병 종식 날짜가 웹에 출력되는지 테스트한다.

Traceability matrix

Functional Requirements	Test Case	High-Level Design	Low-Level Design	
			Method	Class
1.1	Test 1	runCrawler()	loadCountryList()	Crawler
1.2	Test 2	loadCountryList()	toTimeSeries()	
1.3	Test 3	crawl()	saveData()	Covid19_Crawler
1.4	Test 4	toTimeSeriese()	crawl()	
2.1	Test 5	saveData()	runCrawler()	Scheduler
2.2	Test 6	runFitter()	runFitter()	
2.3	Test 7	DataFitter	checkScheduler (log path)	
	Test 8		Country()	Country
2.4	Test 9	countryMap()	Disease()	Disease
3.1	Test 10	continentMap()	loadData()	SIR & SEIR
3.2	Test 11		saveData()	
3.3	Test 12		SIR(), SEIR()	
3.4	Test 13		EulerMethod()	Matrix
	Test 14		lsqcurfit()	
			operator+()	
			operator-()	
			operator*()	
			operator=()	
			makeCofactor()	
			determinant()	
			inverse()	
			transposing()	
			searchData()	countryMap
			layerProps()	continentMap
			WorldMap()	

01

Component diagram

02

03

04

05

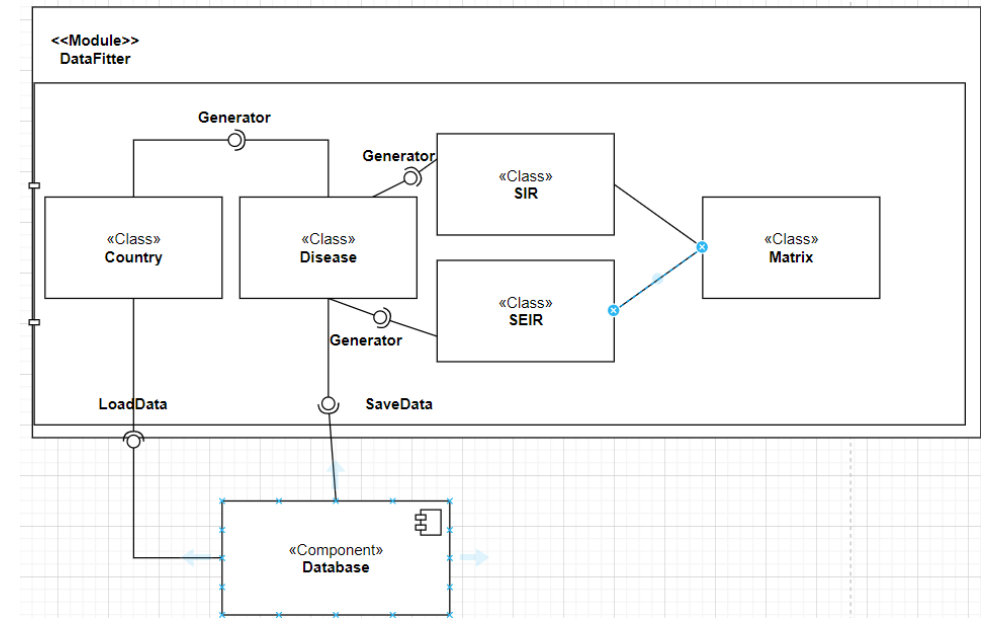
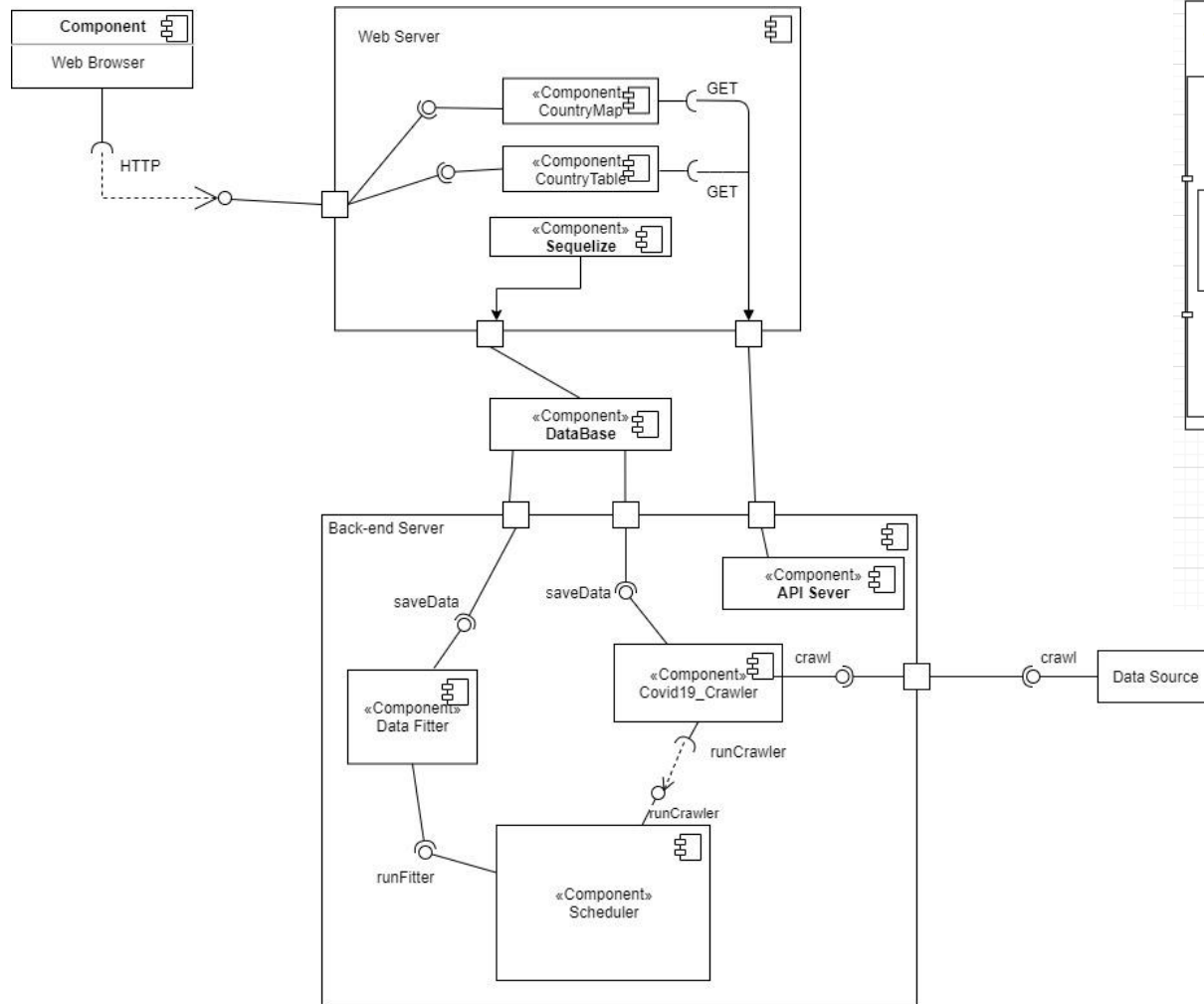
06

07

08

09

10



01

02

Sequence diagram

03

04

05

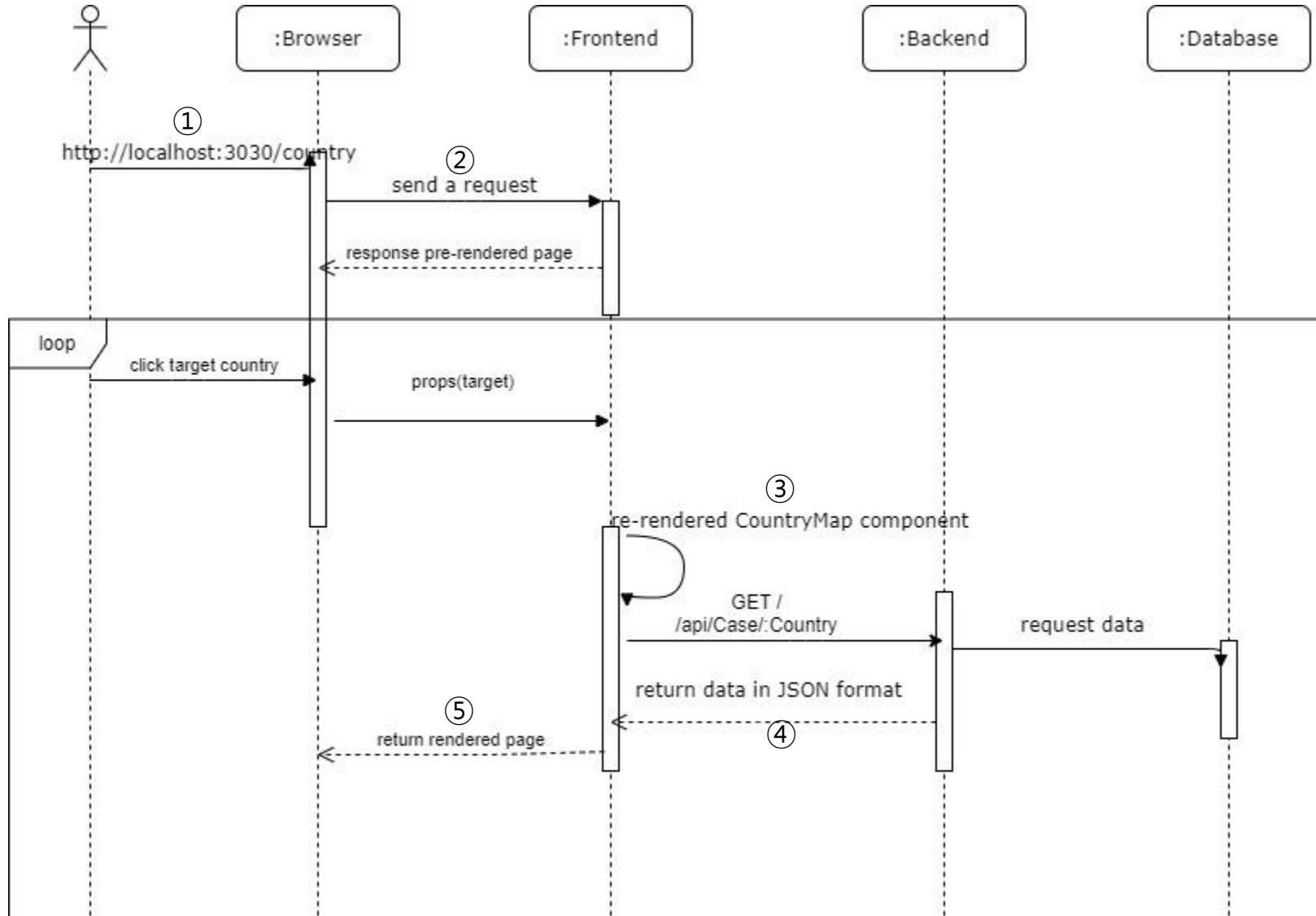
06

07

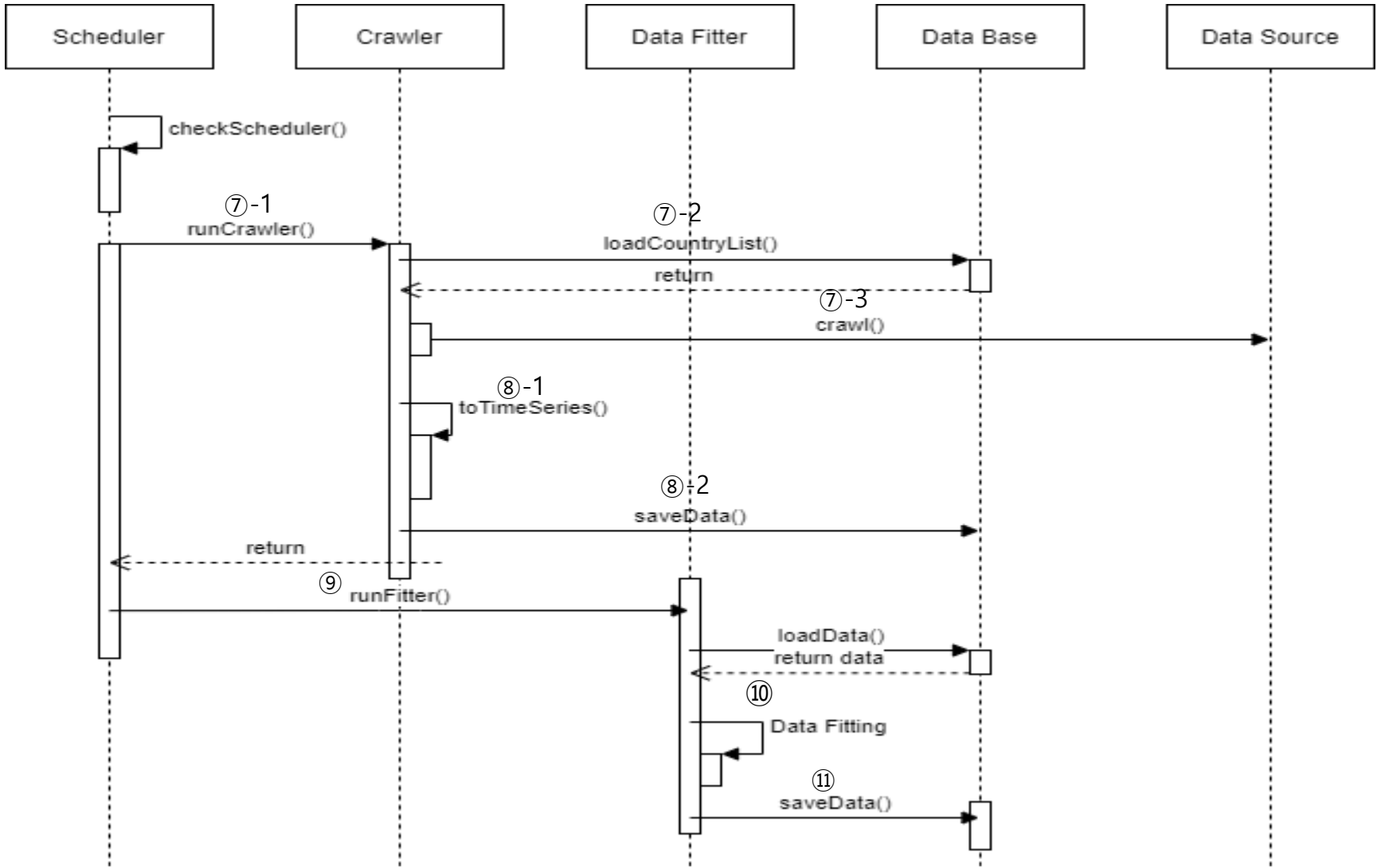
08

09

10



Sequence diagram



04

05

06

07

08

09

10

01

02

03

04

05

06

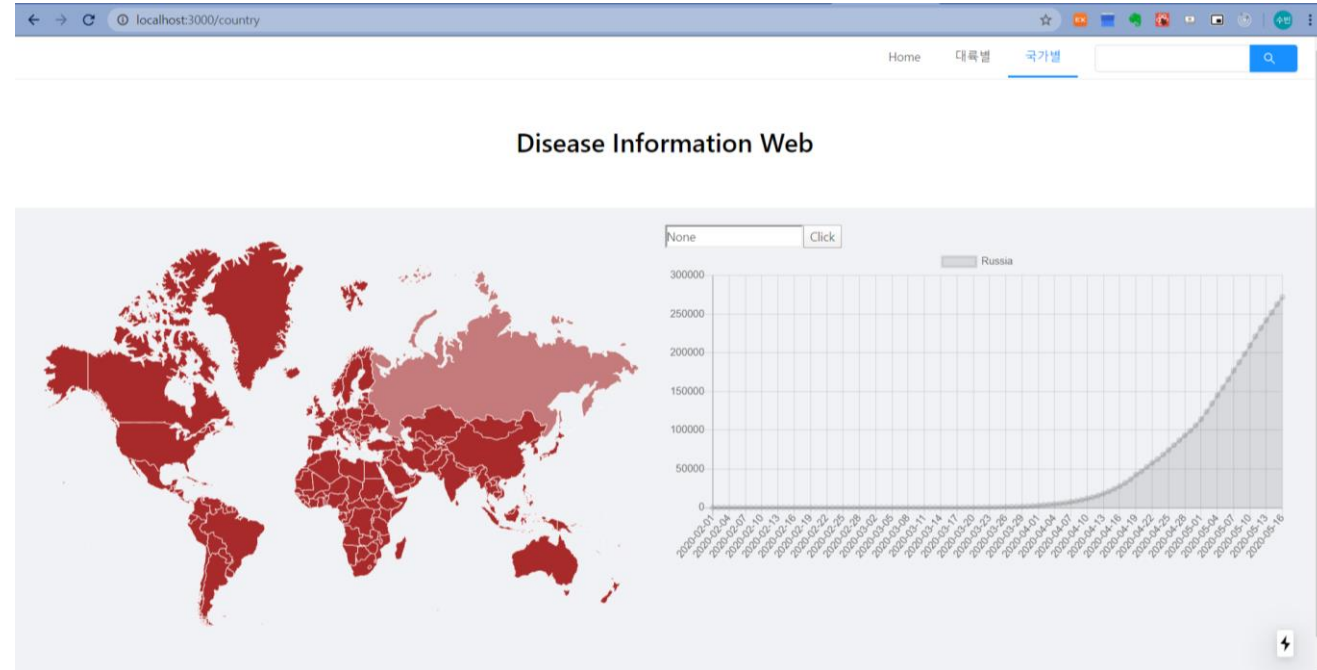
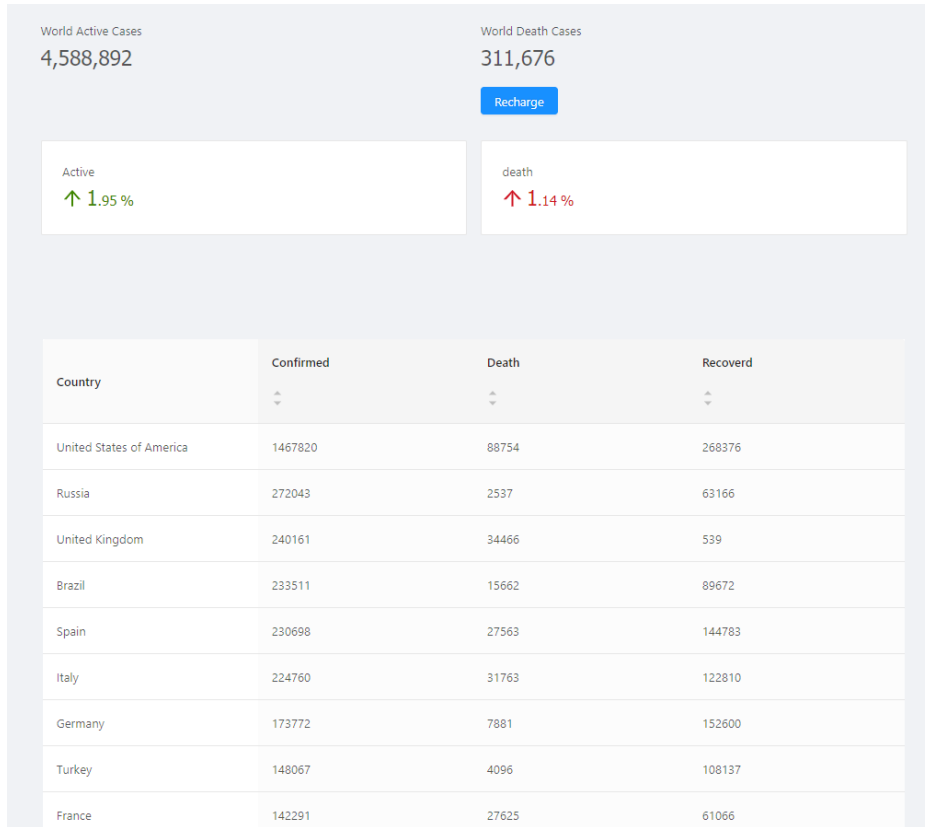
07

08

09

10

1st Iteration



01

02

03

1st Iteration

System Testing Result 7 / 14 (50%)

04

05

06

07

08

09

10

Test#	Test 항목	Pass/Fail Criteria	실패 사유
1	국가별 확진자 수, 사망자 수, 회복 인원을 성공적으로 크롤링하는지 확인한다	(PASS) 데이터 중 None값이 없다.	
		(PASS) 데이터 중 0보다 작은 값이 없다.	
		(PASS) 모든 데이터의 국가이름이 DB에 저장된 국가 이름과 같다.	
2	크롤링한 데이터를 시계열 데이터로 전환하는지 확인한다.	(PASS) 모든 날짜 데이터가 date타입이다. (PASS) 모든 날짜 데이터가 대만 표준시 기준 1일전 값과 같다.	
3	데이터를 DB에 성공적으로 저장하는지 확인한다.	(PASS) 데이터의 개수가 크롤링한 날짜로 DB에 저장된 데이터의 개수와 같다.	
4	스케줄러가 24시간에 한번 실행되는지 확인한다	(FAIL) log가 설정한 시간마다 기록되는지 확인	미구현
5	데이터 베이스에서 불러온 데이터가 모델에 알맞은 형태로 가공되는지 테스트한다.	(FAIL) 데이터 베이스에 접근해데이터를 불러온다	미구현
		(PASS) 데이터를 알맞은 형태로 가공되는지 테스트한다.	
6	ODE(Ordinary Differential Equation) Solver의 결과값이 계산되는지 테스트한다.	(PASS) ODE Solver가 ODE의 해를 구한다.	
7	Matrix Class를 이용 행렬의 기본적인 연산(+, -, *, inverse, transpose, 행렬식)이 되는지 확인한다.	(PASS) 행렬의 덧셈이 가능하다.	
		(PASS) 행렬의 뺄셈이 가능하다.	
		(PASS) 행렬의 곱셈이 가능하다.	
		(PASS) 역행렬을 구할 수 있다.	
		(PASS) 행렬식을 계산가능하다.	
8	데이터를 ODE Solver를 이용해 피팅 후 전파율과 회복율이 계산되는지 테스트한다.	(FAIL) 데이터 피팅을 통해 Parameter계산이 가능하다.	NAN 오류
9	계산된 전파율과 회복율을 ODE Solver에 적용하여전염병 유행 종료시기를 예측할 수 있는지 테스트한다.	(FAIL) 피팅된 Parameter를 통해 유행 종료시기 예측이 가능하다.	미구현
10	계산된 상관 계수가 Success Criteria를 만족하는지 테스트한다.	(FAIL) 상관계수를 계산한 후 Success Criteria에 부합하는지 확인한다.	미구현
11	국가를 검색했을 때 해당 국가의 전염병 데이터가 웹에 출력되는지 확인한다.	(FAIL)국가 검색시 확진자 데이터가 나타난다.	미구현
		(FAIL)미구현으로 해당 국가의 종식날짜는 나타나지 않는다.	
12	국가를 선택했을 때 해당 국가의 전염병 확진자 데이터가 표의 형식으로 웹에 출력되는지 확인한다.	(PASS)확진자 데이터가 정상적으로 나타난다.	
13	국가를 선택했을 때 해당 국가의 전염병 확진자 추이가 그래프 형식으로 웹에 출력되는지 확인한다.	(PASS) 국가 선택시, 날짜별 확진자 추이가 그래프로 나타난다.	
14	국가를 선택했을 때 지도 위에 전염병 종식 날짜가 웹에 출력되는지 확인한다.	(FAIL)미구현으로 종식 날짜가 나타나지 않는다.	미구현

01

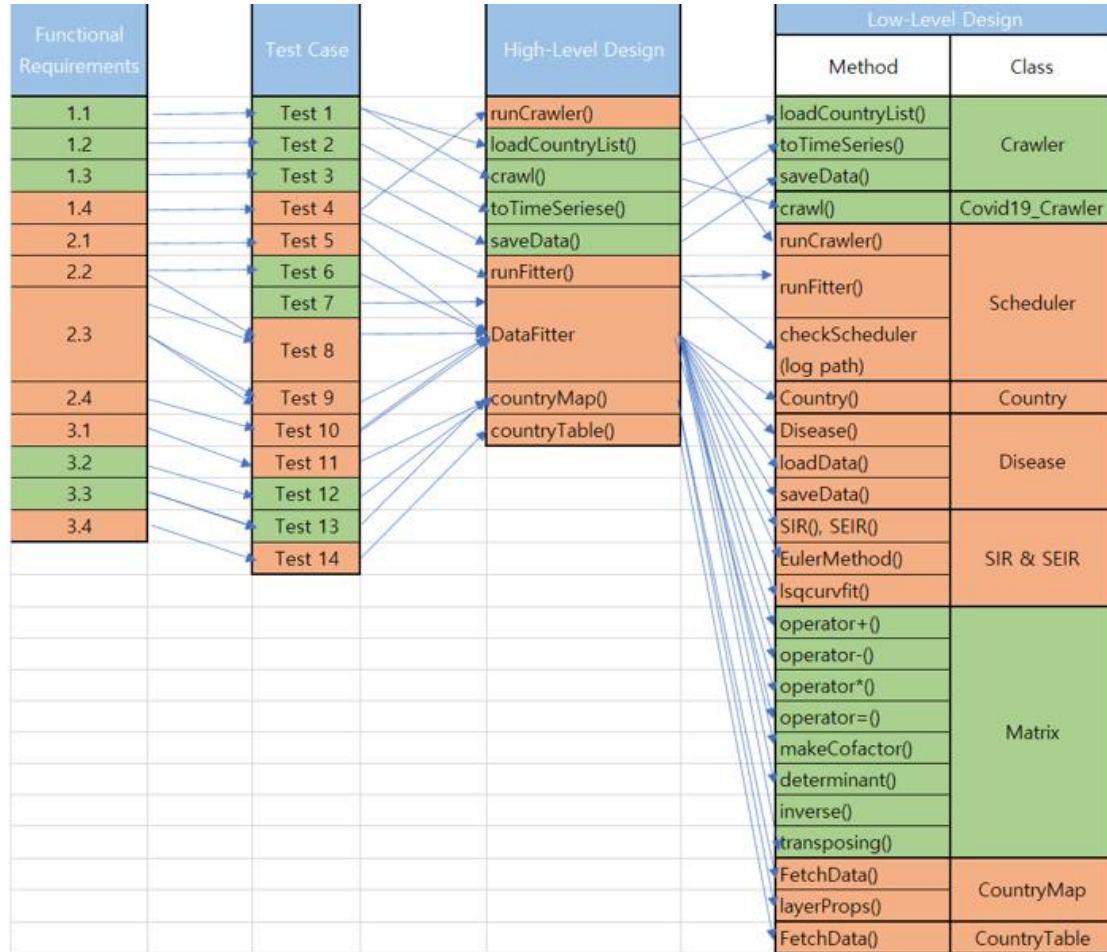
1st Iteration

02

03

System Testing Result 7 / 14 (50%)

04



05

06

07

08

09

10

2nd Iteration

System Testing Result 11 / 14 (78%)

Test#	Test 항목	Pass/Fail Criteria	실패 사유
1	국가별 확진자 수, 사망자 수, 회복 인원을 성공적으로 크롤링하는지 확인한다	(PASS) 데이터 중 None값이 없다.	
		(PASS) 데이터 중 0보다 작은 값이 없다.	
		(PASS) 모든 데이터의 국가이름이 DB에 저장된 국가 이름과 같다.	
2	크롤링한 데이터를 시계열 데이터로 전환하는지 확인한다.	(PASS) 모든 날짜 데이터가 date타입이다. (PASS) 모든 날짜 데이터가 대만 표준시 기준 1일전 값과 같다.	
3	데이터를 DB에 성공적으로 저장하는지 확인한다.	(PASS) 데이터의 개수가 크롤링한 날짜로 DB에 저장된 데이터의 개수와 같다.	
4	스케줄러가 24시간에 한번 실행되는지 확인한다	(PASS) log가 설정한 시간마다 기록되는지 확인	
5	데이터 베이스에서 불러온 데이터가 모델에 알맞은 형태로 가공되는지 테스트한다.	(PASS) 데이터 베이스에 접근해데이터를 불러온다	
		(PASS) 데이터를 알맞은 형태로 가공되는지 테스트한다.	
6	ODE(Ordinary Differential Equation) Solver의 결과값이 계산되는지 테스트한다.	(PASS) ODE Solver가 ODE의 해를 구한다.	
7	Matrix Class를 이용 행렬의 기본적인 연산 (+, -, *, inverse, transpose, 행렬식)이 되는지 확인한다.	(PASS) 행렬의 덧셈이 가능하다.	
		(PASS) 행렬의 뺄셈이 가능하다.	
		(PASS) 행렬의 곱셈이 가능하다.	
		(PASS) 역행렬을 구할 수 있다.	
		(PASS) 행렬식을 계산가능하다.	
8	데이터를 ODE Solver를 이용해 피팅 후 전파율과 회복율이 계산되는지 테스트한다.	(PASS) 데이터 피팅을 통해 Parameter계산이 가능하다.	
9	계산된 전파율과 회복율을 ODE Solver에 적용하여전염병 유행 종료시기를 예측할 수 있는지 테스트한다.	(PASS) 피팅된 Parameter를 통해 유행 종료시기 예측이 가능하다.	
10	계산된 상관 계수가 Success Criteria를 만족하는지 테스트한다.	(FAIL) 상관계수를 계산한 후 Success Criteria에 부합하는지 확인한다.	미구현
11	국가를 검색했을 때 해당 국가의 전염병 데이터가 웹에 출력되는지 확인한다.	(PASS) 국가 검색시 확진자 데이터가 나타난다.	미구현
		(FAIL)미구현으로 해당 국가의 종식날짜는 나타나지 않는다.	
12	국가를 선택했을 때 해당 국가의 전염병 확진자 데이터가 표의 형식으로 웹에 출력되는지 확인한다.	(PASS)확진자 데이터가 정상적으로 나타난다.	
13	국가를 선택했을 때 해당 국가의 전염병 확진자 추이가 그래프 형식으로 웹에 출력되는지 확인한다.	(PASS) 국가 선택시, 날짜별 확진자 추이가 그래프로 나타난다.	
14	국가를 선택했을 때 지도 위에 전염병 종식 날짜가 웹에 출력되는지 확인한다.	(FAIL)미구현으로 종식 날짜가 나타나지 않는다.	미구현

01

2nd Iteration

02

03

System Testing Result 11 / 14 (78%)

04

05

06

07

08

09

10

Functional Requirements	Test Case	High-Level Design	Low-Level Design	
			Method	Class
1.1	Test 1	runCrawler()	loadCountryList()	Crawler
1.2	Test 2	loadCountryList()	toTimeSeries()	
1.3	Test 3	crawl()	saveData()	
1.4	Test 4	toTimeSeries()	crawl()	Covid19_Crawler
2.1	Test 5	saveData()	runCrawler()	Scheduler
2.2	Test 6	runFitter()	runFitter()	
2.3	Test 7	DataFitter	checkScheduler (log path)	
2.4	Test 8	countryMap()	Country()	Country
3.1	Test 9	continentMap()	Disease()	Disease
3.2	Test 10		loadData()	
3.3	Test 11		saveData()	
3.4	Test 12		SIR(), SEIR()	SIR & SEIR
	Test 13		EulerMethod()	
	Test 14		lsqcurfit()	
			operator+()	Matrix
			operator-()	
			operator*()	
			operator=()	
			makeCofactor()	
			determinant()	
			inverse()	
			transposing()	
			searchData()	countryMap
			layerProps()	continentMap
			WorldMap()	

01

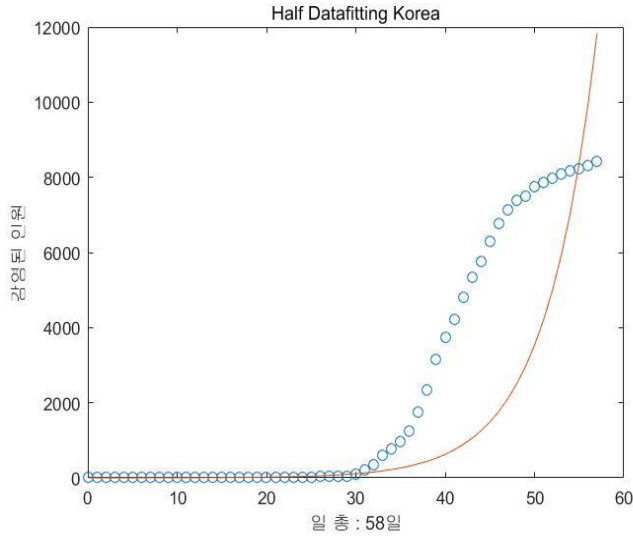
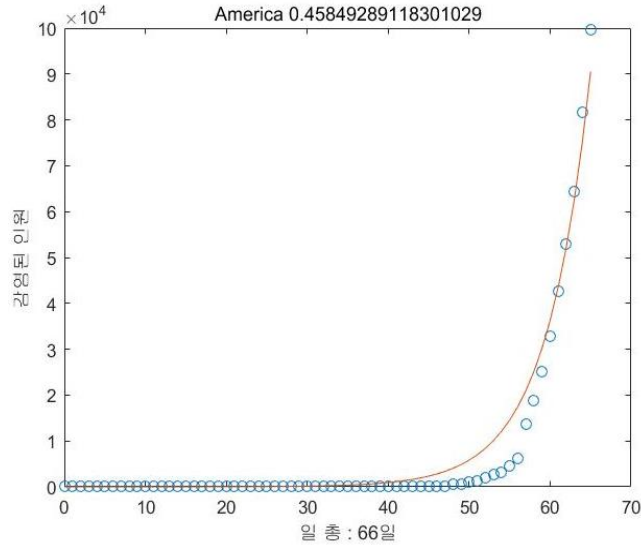
2nd Iteration

02

03

Pass / Fail Criteria

04



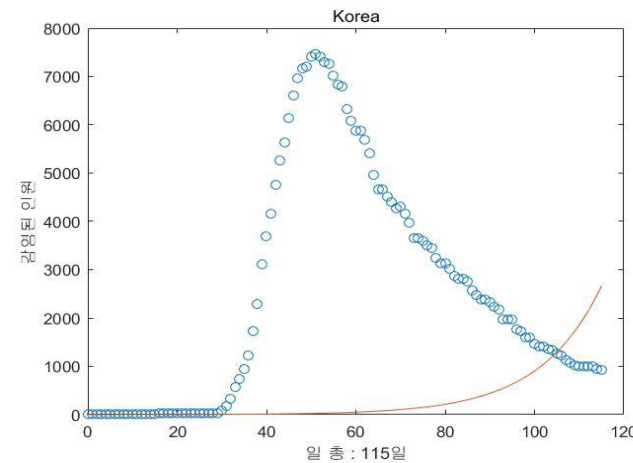
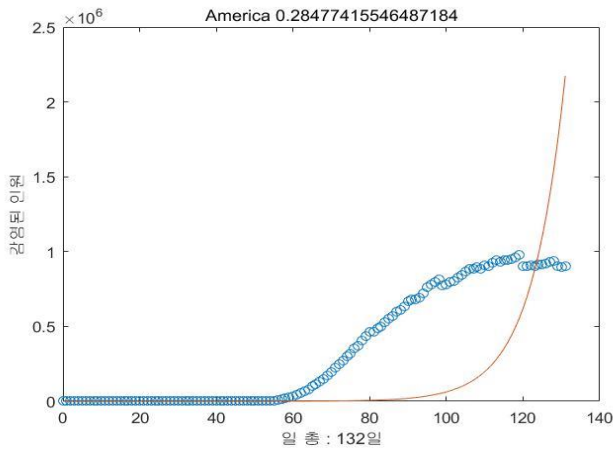
- 분기별 데이터로 예측시 보다 유효한 예측을 보임
- 자연방목 상태일 수록 데이터 유효성이 높음

07

08

09

10



- 전체 데이터로 예측시 방역에 상당한 노력이 들어간 국가일 경우 유효성이 상당히 낮아짐

➔ 한국과 같은 경우 새로운 모델 필요

01

3rd Iteration

02

03

System Testing Result 14 / 14 (100%)

04

05

06

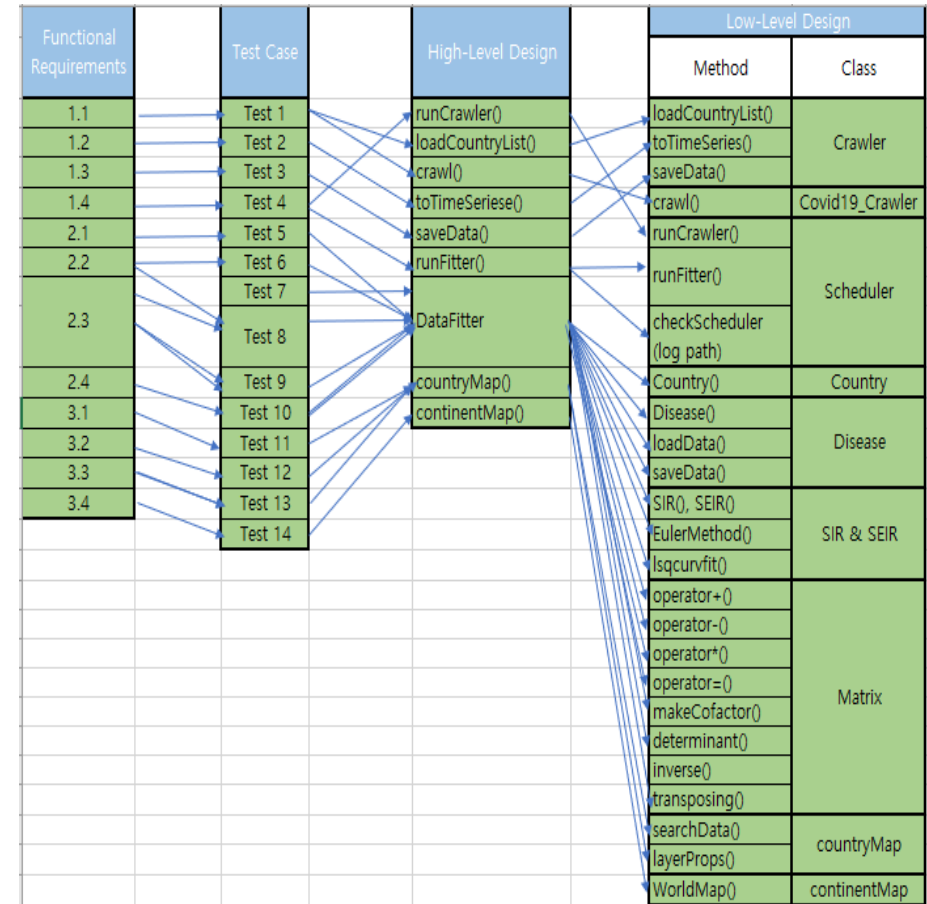
07

08

09

10

Test#	Test 항목	Pass/Fail Criteria
1	국가별 확진자 수, 사망자 수, 회복 인원을 성공적으로 크롤링하는지 확인한다	(PASS) 데이터 중 None값이 없다. (PASS) 데이터 중 0보다 작은 값이 없다. (PASS) 모든 데이터의 국가이름이 DB에 저장된 국가 이름과 같다.
2	크롤링한 데이터를 시계열 데이터로 전환하는지 확인한다.	(PASS) 모든 날짜 데이터가 date타입이다. (PASS) 모든 날짜 데이터가 대만 표준시 기준 1일전 값과 같다.
3	데이터를 DB에 성공적으로 저장하는지 확인한다.	(PASS) 데이터의 개수가 크롤링한 날짜로 DB에 저장된 데이터의 개수와 같다.
4	스케줄러가 24시간에 한번 실행되는지 확인한다	(PASS) log가 설정한 시간마다 기록되는지 확인
5	데이터 베이스에서 불러온 데이터가 모델에 알맞은 형태로 가공되는지 테스트한다.	(PASS) 데이터 베이스에 접근해데이터를 불러온다 (PASS) 데이터를 알맞은 형태로 가공되는지 테스트한다.
6	ODE(Ordinary Differential Equation) Solver의 결과값이 계산되는지 테스트한다.	(PASS) ODE Solver가 ODE의 해를 구한다.
7	Matrix Class를 이용 행렬의 기본적인 연산 (+, -, *, inverse, transpose, 행렬식)이 되는지 확인한다.	(PASS) 행렬의 덧셈이 가능하다. (PASS) 행렬의 뺄셈이 가능하다. (PASS) 행렬의 곱셈이 가능하다. (PASS) 역행렬을 구할 수 있다. (PASS) 행렬식을 계산가능하다.
8	데이터를 ODE Solver를 이용해 피팅 후 전파율과 회복율이 계산되는지 테스트한다.	(PASS) 데이터 피팅을 통해 Parameter계산이 가능하다.
9	계산된 전파율과 회복율을 ODE Solver에 적용하여전염병 유행 종료시기를 예측할 수 있는지 테스트한다.	(PASS) 피팅된 Parameter를 통해 유행 종료시기 예측이 가능하다.
10	계산된 상관 계수가 Success Criteria를 만족하는지 테스트한다.	(PASS) 상관계수를 계산한 후 Success Criteria에 부합하는지 확인한다.
11	국가를 검색했을 때 해당 국가의 전염병 데이터가 웹에 출력되는지 확인한다.	(PASS)국가 검색시 확진자 데이터가 나타난다. (PASS)미구현으로 해당 국가의 종식날짜는 나타나지 않는다.
12	국가를 선택했을 때 해당 국가의 전염병 확진자 데이터가 표의 형식으로 웹에 출력되는지 확인한다.	(PASS)확진자 데이터가 정상적으로 나타난다.
13	국가를 선택했을 때 해당 국가의 전염병 확진자 추이가 그래프 형식으로 웹에 출력되는지 확인한다.	(PASS) 국가 선택시, 날짜별 확진자 추이가 그래프로 나타난다.
14	국가를 선택했을 때 지도 위에 전염병 종식 날짜가 웹에 출력되는지 확인한다.	(PASS)미구현으로 종식 날짜가 나타나지 않는다.



01

3rd Iteration

02

03

실행 화면

04

Disease Information Web

05

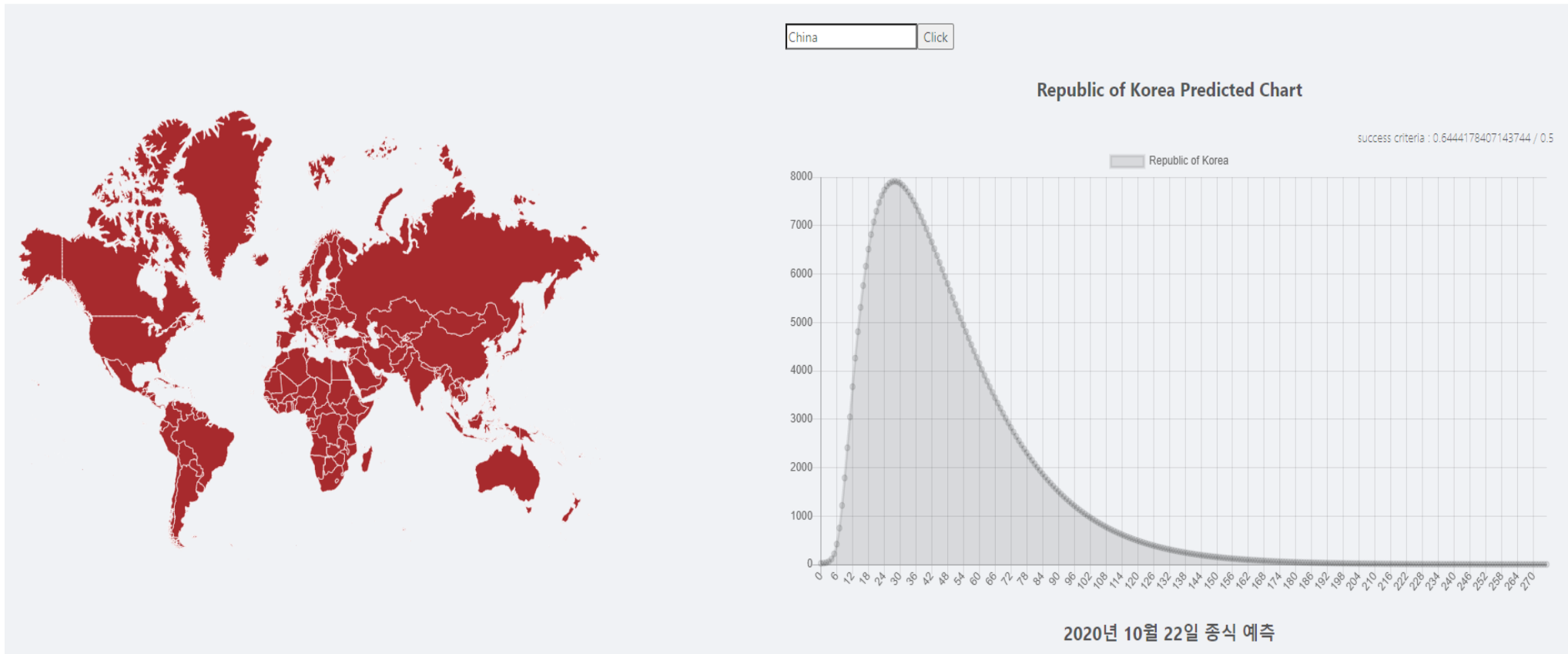
06

07

08

09

10



01

3rd Iteration

02

Pass / Fail Criteria

03

초기 목표 : 결정계수(R^2) 0.5 이상

04

05

2nd Iteration에서 발견한 문제점

06

- 한국의 낮은 유효성 -> 건국대 정은옥 교수님 팀의 모델을 이용해 해결
- 그 외 국가의 유효성 -> 새로운 질병 예측 모델인 SEIQR 모델을 이용해 더 높임

07

08

최종 결과

09

- Covid-19 감염국 : 총 213개국
- 결정계수(R^2) 0.5 이상 달성 : 206개국 (96.7%)
- 최저값 : 0.483(프랑스) 목표치 대비 -3.4%
- 213개국의 평균 결정계수 : 0.67 목표치 대비 34% 초과 달성

10

