Python을 이용한 웹 스크레이핑 실습

1. Python

1. 설치 및 개발환경 구축

1. 통합 개발 환경(Integrated Development Environment, IDE)

- 코딩, 디버깅, 컴파일, 배포 등 프로그램 개발에 관련된 모든 작업을 하나의 프로그램 안에서 처리하는 환경을 제공하는 소프트웨어
  - IDLE (Integrated Development and Learning Environment)
    - Python에서 설치 가능
    - 가장 기본적인 IDE
  - Pycharm
    - JetBrains사에서 개발한 Python 특화 IDE
  - Jupyter Notebook
    - Markdown과 연동하여 Python 코드 블록을 포함한 문서 작성 가능
    - 현재 이 교육자료는 Jupyter Notebook으로 제작
  - Spyder
    - 과학 분야의 프로그래밍을 위해 개발된 IDE
    - Line-by-line 실행이 편리하여 초보자에게 적합하므로 본 교육세션에서 사용

2. 아나콘다(Anaconda)

- Python 및 R을 과학연구 및 기계학습 분야에서 활용하기 편하도록 각종 라이브러리와 개발환경을 설치해주는 소프트웨어
  - 아나콘다를 사용하는 것은 자동차를 살 때 옵션을 선택하는 것과 같다.
  - 즉, 아나콘다 없이 옵션이 없는 순정차(기본 Python)을 사용해도 문제는 없지만 옵션을 선택하는 것이 편리한 것과 같음
- 본 교육세션에서는 아나콘다를 이용하여 Python 설치
  - 아나콘다 다운로드 페이지에서 자신의 운영체제에 맞는 Individual Edition 설치
  - Anaconda를 설치하면 Spyder와 Jupyter Notebook도 자동으로 설치

2. Spyder 기본 사용법

- 스크립트(script): 실제로 코드를 짜는 부분
- 콘솔(console): 코드의 실행결과가 나타나는 부분
- 변수 탐색기(variable explorer): 내가 정의한 변수의 목록과 값을 확인할 수 있는 창.
- 실행과 디버깅(run and debug)
- 한 줄 실행 단축키 설정: F9 → Ctrl + E

In [38]:
```
# Hello World!
print("Hello, World!")
```

```
Hello, World!
```

2. 변수(variables)

1. 변수의 할당(assignment)

- 프로그램이 데이터를 기억하는 방식으로서 컴퓨터의 메모리의 공간에 이름을 붙이는 것으로 음식 재료를 사다가 그릇안에 담아 놓은 것
- 엑셀의 "이름정의"와 유사
- 바뀔 수 없는 리터럴(literal)과 달리 바뀔 수 있기(variable) 때문에 변수만 새롭게 정의하면 해당 변수가 사용된 모든 곳에 변화가 적용.
- Python은 `=` 연산자를 사용하여 변수의 선언(declaration)과 할당(assignment)를 동시에 가능
- input() 함수를 이용하면 사용자로부터 변수를 입력받을 수 있음

2. 변수의 명명(naming)

- 변수의 명명은 프로그래머의 자유이나 서로 다른 변수들끼리 식별될 수 있도록 몇 가지 규칙을 따름
  - 변수의 이름은 영문자와 숫자, 밑줄 문자(_)로 이루어짐
  - 식별자의 중간에 공백이 들어가면 안됨
  - 식별자의 첫 글자는 반드시 영문자 또는 밑줄 문자(_)이어야 한다. 즉, 숫자로 시작할 수 없음
  - 대문자와 소문자는 구별된다. (case-sensitive)
- 또한 미리 정의된 Python 키워드(keyword)는 변수명으로 사용될 수 없다. (e.g. True, False, for, while 등)
- 변수의 이름은 해당 변수의 내용을 잘 설명하도록 지어야 한다. 변수의 이름이 잘 지어져야 읽기 쉬운 프로그램
- 중구난방으로 지어진 변수는 프로그램이 조금만 길어지거나 오랜만에 다시 보게 코드의 가독성을 떨어뜨림
- 명명 규칙(Case Naming Convention)
  - R: lowerCamelCase (e.g. `myPython`)
  - Python: snake_case (e.g. `my_python`)

In [39]:
```
# 변수의 명명 및 할당
my_variable = 10
_my_variable = 20
my_variable_ = 30
```

In [40]:
```
print(my_variable)
print(_my_variable)
print(my_variable_)
```

```
10
20
30
```

3. 연산(operation)

- 수식(expression)이란 피연산자(operand)들과 연산자(operator)들의 조합.
  - 피연산자는 연산의 대상이 되는 것을 의미.
  - 연산자는 어떤 연산을 나타내는 기호를 의미.
  - 컴퓨터는 인간을 위하여 복잡한 계산을 대신 해주지만, 우리가 올바른 수식을 알려주지 않는다면 그 계산 결과는 틀릴것
  - 올바른 수식을 작성하기 위해서는 연산자의 기능과 올바른 순서를 알고 있어야 함

In [41]:
```
# 사칙연산 연산자
print(2+3)
print(2-3)
print(2*3)
print(2/3)
```

```
5
-1
6
0.6666666666666666
```

In [42]:
```
# 나눗셈의 몫과 나머지
p = 10
q = 3
quotient = 10 // 3
remainder = 10 % 3
print(p,"/",q,"의 몫은 ",quotient, ", 나머지는 ", remainder, "입니다.")
```

```
10 / 3 의 몫은  3 , 나머지는  1 입니다.
```

In [43]:
```
# 거듭제곱
print(2 ** 3)
```

```
8
```

In [44]:
```
# 할당
powered = 2 ** 3
print(powered)
```

```
8
```

In [45]:
```
# 논리연산자
print(1 == 1)
print(1 < 0)
```

```
True
False
```

4. 자료형(data type)

- 변수가 음식을 담는 그릇이라면 자료형은 그릇의 모양
  - Python의 자료형에는 대표적으로 정수(int), 실수(float), 불(bool), 문자열(str), 리스트(list), 튜플(tuple), 집합(set), 딕셔너리(dict)가 있음
  - `type()` 함수를 통해 자료형을 확인 가능

In [46]:
```
# 정수
data = 1
print(type(data))
```

```
<class 'int'>
```

In [47]:
```
# 실수
data = 1.2
print(type(data))
```

```
<class 'float'>
```

In [48]:
```
# 불
data = 1 == 1
print(data)
print(type(data))
```

```
True
<class 'bool'>
```

In [49]:
```
# 문자열
```

```
data = "abc"
print(data[0])
print(type(data))
```

```
a
<class 'str'>
```

In [50]:
```
# 리스트
data = [1, 2, 3, 4]
print(data[1])
print(type(data))
```

```
2
<class 'list'>
```

In [51]:
```
# 튜플
data = (1, 2, 3)
print(data[2])
print(type(data))
```

```
3
<class 'tuple'>
```

In [52]:
```
# 집합
data = {1, 2, 3, 3, 2, 1}
print(data)
print(type(data))
```

```
{1, 2, 3}
<class 'set'>
```

In [53]:
```
# 딕셔너리
data = {
    "사과": 1,
    "배": 2,
    "포도": 3
}
print(data["사과"])
print(type(data))
```

```
1
<class 'dict'>
```

5. 조건문(conditional statement)

- 프로그래밍을 하다보면 조건에 따라 서로 다른 명령을 실행해야하는 경우가 있음.
- 프로그래밍 언어에서 선택 구조를 실행할 수 있는 구문을 '조건문(conditional statement)'라고 함.

1. `if` 와 `else

- 조건이 한 개밖에 없는 경우 키워드 'if'와 'else'만으로 조건문을 만들 수 있음.

In [54]:
```
# 합격/불합격 판별기
score = int(input("성적을 입력하시오: "))

if score >= 60:
    print("합격입니다.")
else:
    print("불합격입니다.")
```

```
성적을 입력하시오: 100
합격입니다.
```

In [55]:
```
# 홀짝 구분기
num = int(input("정수를 입력하시오: "))

if num % 2 == 0:
    print("짝수입니다.")
else:
    print("홀수입니다.")
```

```
정수를 입력하시오: 2
짝수입니다.
```

2. elif

- 조건이 두 개 이상인 경우 키워드 'elif'를 사용하여 조건을 추가할 수 있음.

In [56]:
```
# 정수 부호 판별기
num = int(input("정수를 입력하시오: "))

if num > 0:
    print("양수입니다.")
elif num == 0:
    print("0입니다.")
else:
    print("음수입니다")
```

```
정수를 입력하시오: 0
0입니다.
```

3. 중첩 if 문(nested if statement)

- if문 안에 다른 if문이 들어가는 것을 중첩 if문이라고 함

In [57]:
```
# 중첩 if문
num = int(input("정수를 입력하시오: "))

if num >= 0:
    if num == 0:
        print("0입니다.")
    else:
        print("양수입니다.")
else:
    print("음수입니다.")
```

```
정수를 입력하시오: -2
음수입니다.
```

6. 반복문(repetition statetment)

- 우리가 컴퓨터를 쓰는 이유는 인간의 수고를 덜기 위한 것.
- 동일한 작업은 반복 구조를 이용하여 프로그래밍하는 것이 효율적.
- 프로그래밍에서 반복적인 작업을 실행할 수 있도록 하는 구문을 '반복문(repetition statement)'이라고 함
- Python에는 2가지 종류의 반복이 있음:
  - 횟수 제어 반복(for 문): 정해진 횟수만큼 반복.
  - 조건 제어 반복(while 문): 특정한 조건이 만족되면 계속 반복.

1. for 문 (횟수 제어 반복)

- 많은 프로그래밍 언어에서 for문을 이용하여 횟수 제어 반복을 제공.

- 반복 횟수를 알고 있을 때 사용.

In [58]:
```
# for문 예시
numbers = ["하나", "둘", "셋", "넷", "다섯"]
for number in numbers:
    print(number+"!")
```

```
하나!
둘!
셋!
넷!
다섯!
```

- `range()` 는 옵션에 따라 특정한 규칙대로 숫자들을 반환하는 함수.
- `range(시작값, 종료값+1, 증분값)` 의 형태로 사용하며, 기본값으로서 시작값은 0, 증분값은 1.

In [59]:
```
# range() 함수 1
print(range(5))
print(list(range(5)))
```

```
range(0, 5)
[0, 1, 2, 3, 4]
```

In [60]:
```
# range() 함수 2
print(range(1, 6))
print(list(range(1, 6)))
```

```
range(1, 6)
[1, 2, 3, 4, 5]
```

In [61]:
```
# range() 함수 3
print(range(0, 10, 2))
print(list(range(0, 10, 2)))
```

```
range(0, 10, 2)
[0, 2, 4, 6, 8]
```

- 이하 두 코드는 동일한 기능을 수행

In [62]:
```
# 리스트의 인덱스로서의 range() 함수 1
for index in range(len(numbers)):
    print(numbers[index])
```

```
하나
둘
셋
넷
다섯
```

In [63]:
```
# 리스트의 인덱스로서의 range() 함수 2
for index in [0, 1, 2, 3, 4]:
    print(numbers[index])
```

```
하나
둘
셋
넷
다섯
```

2. while 문 (조건 제어 반복)

- Python에서는 조건 제어 반복을 위하여 while문을 제공.
- 반복해야하는 조건을 알고 있을 때 사용.

In [64]:
```
# while문 예시
counter = 1
while counter <= 5:
    print(f"{counter}번!")
    counter = counter + 1
```

```
1번!
2번!
3번!
4번!
5번!
```

7. 함수(function)

1. 추상화(abstraction)와 문제 분해(decomposition)

- 프로그래밍은 문제의 해결과정이며, 추상화와 문제 분해는 효과적인 문제 해결 방법.
- 어떠한 것이 "어떻게"보다는 "무엇을"하는지에 초점을 맞추는 방법을 추상화라고 함.
- 큰 문제를 여러가지의 작은 문제로 쪼개어서 생각하는 것을 문제 분해라고 함.

2. 함수의 정의

- 함수는 일을 수행하는 코드의 덩어리로서, 큰 프로그램을 구축하기 위한 추상화와 문제 분해의 유용한 도구.
- 함수는 입력을 받아서 정해진 규칙에 따라 출력을 내보내는 블랙박스로 생각할 수 있음.
- 함수는 한 번 정의해놓으면 매번 똑같은 동작을 반복하기 때문에 편리.
- 함수는 정의되더라도 호출되기 전까지는 실행되지 않음.
- 함수를 정의하는 방법
  - 함수 정의는 키워드 'def'로 시작.
  - 공백을 한 칸 띄우고 함수의 이름을 입력한다. 함수의 이름에는 공백이 들어갈 수 없음.
  - 함수의 이름 바로 뒤에 소괄호를 입력하고, 인자(arguement)가 있으면 입력.
  - 바로 뒤에 콜론(:)을 입력한 뒤, 다음 줄부터 함수 내용에 해당하는 블록을 입력하며, 블록은 반드시 들여쓰기 되어야 함.

In [65]:
```
# 함수의 정의 예시
def print_my_address():
    print("양지성: ", "대전광역시 유성구")
    print("John Doe: ", "Chapel Hill, North Carolina")

# 함수의 호출 예시
print_my_address()
```

```
양지성:  대전광역시 유성구
John Doe:  Chapel Hill, North Carolina
```

3. 인수(arguement)와 매개변수(parameter)

- 외부에서 함수에 전달하는 값을 인수라고 함. 정의하기에 따라 함수는 인수를 갖지 않을 수도 있음.
- 함수가 정의될 때 인수가 입력되는 변수를 매개변수라고 함.
- 인수는 두 개 이상이 될 수도 있음.

In [66]:
```
# name, address는 매개변수에 해당한다.
def print_my_address(name, address):
    print(name, address)
```

```
# "홍길동", "서울특별시 종로구"는 인수에 해당한다.
print_my_address("양지성", "대전광역시 서구")
```

```
양지성 대전광역시 서구
```

8. 라이브러리(library)

1. 모듈(module)과 패키지(package)

- 프로그래밍이 무에서 유를 창조하는 매우 유용한 활동이다. 하지만 그렇다고 해서 다른 사람이 이미 만들어놓은 것이 있는데 굳이 새롭게 처음부터 만들 필요는 없을 것.
- 모듈이란 이런 비효율의 문제를 피하기 위해 전역변수, 함수 등을 모아놓은 프로그램의 부품과 같은 것으로서 대부분의 프로그래밍 언어에서 지원.
- 패키지는 여러 모듈들이 구조화되어 모여있는 것을 의미.

2. 라이브러리 불러오기

- 라이브러리는 모듈과 패키지로 구성.
- Python에서는 다양한 목적을 위한 라이브러리를 지원하고 있고, 키워드 'from'과 'import'를 활용하여 이루어짐.
- 라이브러리를 불러오는 방법은 크게 다음과 같음:
  - import 라이브러리
  - import 라이브러리 as 별명
  - from 라이브러리 import 모듈

3. `random` 라이브러리

- `random` 은 숫자를 무작위로 생성할 때 유용하게 사용할 수 있는 라이브러리.

In [67]:
```
# import 라이브러리
import random
random.randint(0, 10)
```

Out[67]: 5

In [68]:
```
# import 라이브러리 as 별명
import random as rd
rd.randint(0, 10)
```

Out[68]: 7

In [69]:
```
# from 라이브러리 import 모듈
from random import randint
randint(0, 10)
```

Out[69]: 5

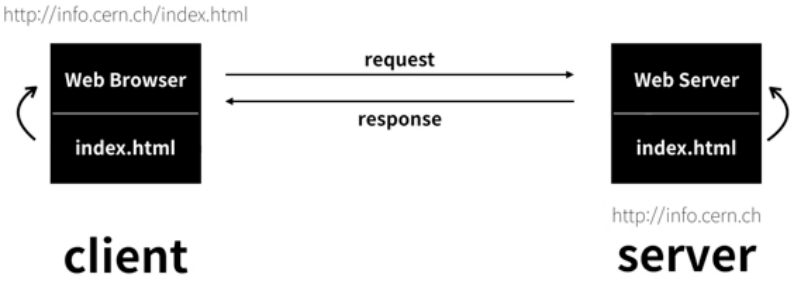
2. 웹 스크레이핑(Web Scraping)

1. 웹(Web)

- World Wide Web (WWW, W3)은 인터넷에 연결된 컴퓨터를 통해 사람들이 정보를 공유할 수 있는 세계적 정보 공간
- 간단하게 줄여서 웹(the Web)이라고 부름

2. HTTP (HyperText Transfer Protocol)

- HTTP는 웹 상에서 정보를 주고 받을 수 있는 프로토콜(규약).
- 클라이언트(client)와 서버(server) 사이에 이루어지는 요청(request)/응답(response)로 이루어짐.
  - 클라이언트는 서비스를 사용하는 사용자 혹은 사용자의 단말기를 의미하며 보통 웹 브라우저를 의미(e.g. Chrome, Internet Explorer 등)
  - 서버는 클라이언트에서 웹 페이지에 접근하려고 요청(request)하면 서버에서는 요청 받은 웹 페이지를 클라이언트 쪽으로 보내줌(response)



3. 웹 개발의 주춧돌(The Cornerstones of Web Development)

- 웹 개발에는 HTML, CSS, JavaScript의 3가지 핵심 기술이 있음 (사이트 참조)
  - HTML(Hypertext Mark-up Language)
    - 웹 페이지의 구조(structure)를 기술하며 요소(element)와 속성(attribute)으로 구성
  - CSS (Cascading Style Sheets):
    - HTML의 요소가 브라우저를 통해 시연(display)되는 방식(미적 요소)를 정의
  - JavaScript:
    - HTML과 CSS에 의해 구성된 웹 페이지를 작동하게 하는 동적(dynamic)인 요소를 정의

> HTML과 CSS는 마크업 언어(mark-up language)이고, JavaScript는 프로그래밍 언어(programming language).

1. HTML

- HTML의 각 요소들은 서로간의 관계에 따라 조상 요소, 자손 요소, 부모 요소, 자식 요소가 될 수 있다.
  - 한 요소의 모든 하위 요소를 자손 요소, 한 요소의 **모든 상위 요소를 조상 요소**라고 한다.
  - 한 요소의 모든 직속 하위 요소를 자식 요소, 한 요소의 **모든 직속 상위 요소를 부모 요소**라고 한다.

In [93]:
```
from distutils.sysconfig import get_python_lib
```

In [94]:
```
from IPython.core.display import HTML
import codecs

with codecs.open("./src/const_html.html", "r", "utf-8") as f:
    html = f.read()
    print(html)
    f.close()
```

```
<!DOCTYPE HTML>
<html>
    <head>
        <meta charset="utf-8">
        <title>대한민국헌법 (HTML)</title>
    </head>
    <body>
        <h1>
            <a href="https://www.law.go.kr/lsEfInfoP.do?lsiSeq=61603#", target="_blank">대한민국헌법
</a>
        </h1>

            <h2>전문</h2>

                <button type="button" class="collapsible">펼치기/접기</button>
                <p class="preface">
                유구한 역사와 전통에 빛나는 우리 대한국민은 3·1운동으로 건립된 대한민국임시정부의
                법통과 불의에 항거한 4·19민주이념을 계승하고, 조국의 민주개혁과 평화적 통일의 사명에
                입각하여 정의·인도와 동포애로써 민족의 단결을 공고히 하고, 모든 사회적 폐습과 불의를
타파하며,
                자율과 조화를 바탕으로 자유민주적 기본질서를 더욱 확고히 하여 정치·경제·사회·문화
의
                모든 영역에 있어서 각인의 기회를 균등히 하고, 능력을 최고도로 발휘하게 하며,
                자유와 권리에 따르는 책임과 의무를 완수하게 하여, 안으로는 국민생활의 균등한 향상을
기하고
                밖으로는 항구적인 세계평화와 인류공영에 이바지함으로써 우리들과 우리들의 자손의 안전
과
                자유와 행복을 영원히 확보할 것을 다짐하면서 1948년 7월 12일에 제정되고 8차에 걸쳐
                개정된 헌법을 이제 국회의 의결을 거쳐 국민투표에 의하여 개정한다.
                </p>

            <h2>제1장 총강</h2>

                <p>
                    <div>제1조</div>
                    <div>
                        <ol>
                            <li>대한민국은 민주공화국이다.</li>
                            <li>대한민국의 주권은 국민에게 있고, 모든 권력은 국민으로부터 나온다.</li
>
                        </ol>
                    </div>
                </p>

                <p>
                    <div>제2조</div>
                    <div>
                        <ol>
                            <li>대한민국의 국민이 되는 요건은 법률로 정한다.</li>
                            <li>국가는 법률이 정하는 바에 의하여 재외국민을 보호할 의무를 진다.</li>
                        </ol>
```

```
                    </div>
                </p>

                <p>
                    <div><div>(이하 생략)</div>
                </p>


            <h2>제2장 국민의 권리와 의무</h2>

                <p>
                    <div>제10조</div>
                    <div>
                    모든 국민은 인간으로서의 존엄과 가치를 가지며, 행복을 추구할 권리를 가진다.
                    국가는 개인이 가지는 불가침의 기본적 인권을 확인하고 이를 보장할 의무를 진다.
                    </div>
                </p>

                <p>
                    <div>제11조</div>
                    <div>
                        <ol>
                            <li>모든 국민은 법 앞에 평등하다. 누구든지 성별·종교 또는 사회적 신분에
의하여
                            정치적·경제적·사회적·문화적 생활의 모든 영역에 있어서 차별을 받지 아니
한다.</li>
                            <li>사회적 특수계급의 제도는 인정되지 아니하며, 어떠한 형태로도 이를 창설
할 수 없다.</li>
                            <li>훈장등의 영전은 이를 받은 자에게만 효력이 있고, 어떠한 특권도 이에 따
르지 아니한다.</li>
                        </ol>
                    </div>
                </p>

                <p>
                    <div>(이하 생략)</div>
                </p>

    </body>
</html>
```

2. CSS

In [95]:
```
with codecs.open("./src/const.css", "r", "utf-8") as f:
    css = f.read()
    print(css)
    f.close()
```

```
h1, h2{
  text-align: center;
}

ol > li {
    font-style: italic;
}

ol > li:nth-child(1) {
  color: blue;
}

.collapsible {
  background-color: #eee;
  color: #444;
  cursor: pointer;
  padding: 12px;
  border: none;
  outline: none;
  font-size: 15px;
}
```

```
.active, .collapsible:hover {
  background-color: #ccc;
}

.preface {
  padding: 0 18px;
  display: block;
  overflow: hidden;
}
```

3. JavaScript

In [96]:
```
with codecs.open("./src/const.js", "r", "utf-8") as f:
    js = f.read()
    print(js)
    f.close()
```

```
alert("반갑습니다, 어서오세요!")

var coll = document.getElementsByClassName("collapsible");
var i;

for (i = 0; i < coll.length; i++) {
    coll[i].addEventListener("click", function() {
        this.classList.toggle("active");
        var content = this.nextElementSibling;
        if (content.style.display == "block") {
            content.style.display = "none";
        } else {
            content.style.display = "block";
        }
    });
}
```

4. 대한민국 헌법 웹페이지 스크레이핑 및 분석 실습

- 데이터 과학 연구 프로세스는 일반적으로 다음과 같은 과정을 거침
  1. 데이터 수집: 웹에서 헌법 관련 웹 페이지 소스를 불러와서 헌법 전문 및 조문 추출
  2. 데이터 전처리: 추출된 텍스트에서 불용어 및 구두점 제거 후 토큰화
  3. 데이터 분석: 빈도분석
  4. 데이터 시각화: 워드 클라우드 만들기
  5. 데이터 해석: 우리나라 헌법에 내재된 의미 해석하기
- 본 교육세션에서는 Python을 통해 위키문헌에서 제공하는 대한민국헌법 (제10호) 항목)에서 헌법 텍스트를 수집한뒤 전처리, 분석, 시각화하, 해석하는 일련의 과정을 속성으로 체험

1. 데이터 수집

- 웹 상에는 수많은 데이터가 존재하지만 우리가 원하는 정보와 구조로 항상 제공되지 않음
- 따라서 우리의 분석 목적에 따라 직접 맞춤형 데이터를 수집해야하는 경우 발생
- 웹 스크레이핑(web scraping)이란 웹 페이지를 그대로 가져와서 데이터를 추출해내는 것을 의미하며, 반복적이고 자동화된 스크레이핑을 크롤링(crawling)이라고 함
- 웹크롤링을 하기 위해서는 웹 페이지 내의 HTML 요소에 접근해야한다.
- 개발자 도구를 열어 DOM 구조를 확인해보면 간단해 보이는 웹페이지에도 수많은 요소들이 존재한다는 것을 알 수 있다.

In [97]:
```
import requests
import pickle
import pandas as pd
```

```
pickled = False

if pickled == False:
    # 페이지 소스 가져오기
    result = []

    # 접속정보
    headers = {"User-Agent": W
               """Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Ch
    }

    # 요청을 보내는 URL
    url = "https://ko.wikisource.org/wiki/%EB%8C%80%ED%95%9C%EB%AF%BC%EA%B5%AD%ED%97%8C%EB%B2%95_(%EC

    # 서버에 요청을 보내고 응답 받기
    response = requests.get(url, headers=headers)
    html = response.text

    # 불러온 파일 저장
    with open("const_html.pickle", "wb") as f:
        pickle.dump(html, f)
        f.close()
else:
    html = pd.read_pickle('const_html.pickle')
```

In [98]:

```
# HTML 확인 및 파싱
from bs4 import BeautifulSoup

soup = BeautifulSoup(html, "html.parser")
```

- CSS 선택자는 우리가 수많은 요소 중 우리가 원하는 것을 특정(specify)할 수 있도록 도와준다.

  - CSS 선택자 기본 서식
    - *: 모든 요소 선택
    - 요소이름: 요소 기반 선택
    - 요소이름.클래스이름: 클래스 기반 선택
    - #id: id 속성 기반 선택
  - CSS 선택자 요소 관계 지정 서식
    - 요소이름, 요소이름: 쉼표로 구분된 여러개의 선택자 모두 선택
      - 요소이름 요소이름: 앞 선택자의 후손 중 뒤 선택자에 해당하는 것 모두 선택(하위요소 다)
      - 요소이름 > 요소이름: 앞 선택자의 자손 중 뒤 선택자에 해당하는 것 모두 선택(직속 하위요소만)
      - 선택자1 + 선택자2: 같은 계층에서 바로 뒤에 있는 요소 선택(선택자1 제외)
      - 선택자1 ~ 선택자2: 선택자1부터 선택자2까지의 요소 모두 선택(선택자1 제외)
  - CSS 선택자 속성 지정 서식
    - 요소[속성]: 해당 속성을 가진 요소 선택
    - 요소[속성="val"]: 속성의 값이 정확하게 'val'과 일치하는 요소 선택
    - 요소[속성|="val"]: 속성의 값이 정확하게 'val'이거나 val-'로 시작하는 요소 선택
    - 요소[속성^="val"]: 속성의 값이 val'로 시작하면 선택
    - 요소[속성$="val"]: 속성의 값이 val'로 끝나면 선택
    - 요소[속성*="val"]: 속성의 'val'을 포함하고 있다면 선택
    - 요소[속성~="val"]: 속성의 값에 'val'이 포함되는 요소(공백으로 분리된 값이 일치해야 함)
  - CSS 선택자 위치 또는 상태를 지정 서식

- 요소:root: 루트 요소
- 요소:nth-child(n): n번째 자식 요소
- 요소:nth-last-child(n): 뒤에서부터 n번째 자식 요소
- 요소:nth-of-type(n): n번째 해당 종류의 요소(BeautifulSoup에서 유일하게 지원)
- 요소:first-child: 첫번째 자식요소
- 요소:last-child: 마지막 자식요소
- 요소:first-of-type: 첫 번째 해당 종류의 요소
- 요소:last-of-type: 마지막 해당 종류의 요소
- 요소:only-child: 자식으로 유일한 요소
- 요소:only-of-type: 자식으로 유일한 종류의 요소
- 요소:empty: 내용이 없는 요소
- 요소:lang(code): 특정 언어로 code를 지정한 요소
- 요소:not(s): s이외의 요소
- 요소:enabled: 활성화된 UI 요소
- 요소:disabled: 비활성화된 UI 요소
- 요소:checked: 체크되어 있는 UI 요소 선택

In [99]:
```
# CSS 선택자 (전문)
css_preface = "div.mw-parser-output blockquote"
```

In [100]:
```
# 전문 텍스트 추출
preface = soup.select(css_preface)
preface_text = [preface[0].string]
print(preface_text)
```

```
['유구한 역사와 전통에 빛나는 우리 대한국민은 3·1운동으로 건립된 대한민국임시정부의 법통과 불의에 항거한 4·19민주이념을 계승하고, 조국의 민주개혁과 평화적 통일의 사명에 입각하여 정의·인도와 동포애로써 민족의 단결을 공고히 하고, 모든 사회적 폐습과 불의를 타파하며, 자율과 조화를 바탕으로 자유민주적 기본질서를 더욱 확고히 하여 정치·경제·사회·문화의 모든 영역에 있어서 각인의 기회를 균등히 하고, 능력을 최고도로 발휘하게 하며, 자유와 권리에 따르는 책임과 의무를 완수하게 하여, 안으로는 국민생활의 균등한 향상을 기하고 밖으로는 항구적인 세계평화와 인류공영에 이바지함으로써 우리들과 우리들의 자손의 안전과 자유와 행복을 영원히 확보할 것을 다짐하면서 1948년 7월 12일에 제정되고 8차에 걸쳐 개정된 헌법을 이제 국회의 의결을 거쳐 국민투표에 의하여 개정한다.']
```

In [101]:
```
# CSS 선택자 (조문-조)
css_provision_li = lambda x: fr"""span[id="{str(x)}"]"""
```

In [102]:
```
# 조문-조 텍스트 추출
provision_li_text = []
for i in range(1, 131):
    provision_li_text.append(soup.select_one(css_provision_li(i)).parent.text)

print(provision_li_text[:20])
```

```
['제1조 ① 대한민국은 민주공화국이다.', '제2조 ① 대한민국의 국민이 되는 요건은 법률로 정한다.', '제3조 대한민국의 영토는 한반도와 그 부속도서로 한다.', '제4조 대한민국은 통일을 지향하며, 자유민주적 기본질서에 입각한 평화적 통일 정책을 수립하고 이를 추진한다.', '제5조 ① 대한민국은 국제평화의 유지에 노력하고 침략적 전쟁을 부인한다.', '제6조 ① 헌법에 의하여 체결·공포된 조약과 일반적으로 승인된 국제법규는 국내법과 같은 효력을 가진다.', '제7조 ① 공무원은 국민전체에 대한 봉사자이며, 국민에 대하여 책임을 진다.', '제8조 ① 정당의 설립은 자유이며, 복수정당제는 보장된다.', '제9조 국가는 전통문화의 계승·발전과 민족문화의 창달에 노력하여야 한다.', '제10조 모든 국민은 인간으로서의 존엄과 가치를 가지며, 행복을 추구할 권리를 가진다. 국가는 개인이 가지는 불가침의 기본적 인권을 확인하고 이를 보장할 의무를 진다.', '제11조 ① 모든 국민은 법 앞에 평등하다. 누구든지 성별·종교 또는 사회적 신분에 의하여 정치적·경제적·사회적·문화적 생활의 모든 영역에 있어서 차별을 받지 아니한다.', '제12조 ① 모든 국민은 신체의 자유를 가진다. 누구든지 법률에 의하지 아니하고는 체포·구속·압수·수색 또는 심문을 받지 아니하며, 법률과 적법한 절차에 의하지 아니하고는 처벌·보안처분 또는 강제노역을 받지 아니한다.', '제13조 ① 모든 국민은 행위시의 법률에 의하여 범죄를 구성하지 아니하는 행위로 소추되지 아니하며, 동일한 범죄에 대하여 거듭 처벌받지 아니한다.', '제14조 모든 국민은 거주·이전의 자유를 가진다.', '제15조 모든 국민
```

```
은 직업선택의 자유를 가진다.', '제16조 모든 국민은 주거의 자유를 침해받지 아니한다. 주거에 대한 압수나 수색을 할 때에는 검사의 신청에 의하여 법관이 발부한 영장을 제시하여야 한다.', '제17조 모든 국민은 사생활의 비밀과 자유를 침해받지 아니한다.', '제18조 모든 국민은 통신의 비밀을 침해받지 아니한다.', '제19조 모든 국민은 양심의 자유를 가진다.', '제20조 ① 모든 국민은 종교의 자유를 가진다.']
```

In [103]:
```
# CSS 선택자 (조문-항)
css_provision_dd = "div.mw-parser-output dd"
```

In [104]:
```
# 조문-항 텍스트 추출
provision_dd = soup.select(css_provision_dd)
provision_dd_text = list(map(lambda x: x.text, provision_dd))
print(provision_dd_text[:20])
```

```
['② 대한민국의 주권은 국민에게 있고, 모든 권력은 국민으로부터 나온다.', '② 국가는 법률이 정하는 바에 의하여 재외국민을 보호할 의무를 진다.', '② 국군은 국가의 안전보장과 국토방위의 신성한 의무를 수행함을 사명으로 하며, 그 정치적 중립성은 준수된다.', '② 외국인은 국제법과 조약이 정하는 바에 의하여 그 지위가 보장된다.', '② 공무원의 신분과 정치적 중립성은 법률이 정하는 바에 의하여 보장된다.', '② 정당은 그 목적·조직과 활동이 민주적이어야 하며, 국민의 정치적 의사형성에 참여하는데 필요한 조직을 가져야 한다.', '③ 정당은 법률이 정하는 바에 의하여 국가의 보호를 받으며, 국가는 법률이 정하는 바에 의하여 정당운영에 필요한 자금을 보조할 수 있다.', '④ 정당의 목적이나 활동이 민주적 기본질서에 위배될 때에는 정부는 헌법재판소에 그 해산을 제소할 수 있고, 정당은 헌법재판소의 심판에 의하여 해산된다.', '② 사회적 특수계급의 제도는 인정되지 아니하며, 어떠한 형태로도 이를 창설할 수 없다.', '③ 훈장등의 영전은 이를 받은 자에게만 효력이 있고, 어떠한 특권도 이에 따르지 아니한다.', '② 모든 국민은 고문을 받지 아니하며, 형사상 자기에게 불리한 진술을 강요당하지 아니한다.', '③ 체포·구속·압수 또는 수색을 할 때에는 적법한 절차에 따라 검사의 신청에 의하여 법관이 발부한 영장을 제시하여야 한다. 다만, 현행범인인 경우와 장기 3년 이상의 형에 해당하는 죄를 범하고 도피 또는 증거인멸의 염려가 있을 때에는 사후에 영장을 청구할 수 있다.', '④ 누구든지 체포 또는 구속을 당한 때에는 즉시 변호인의 조력을 받을 권리를 가진다. 다만, 형사피고인이 스스로 변호인을 구할 수 없을 때에는 법률이 정하는 바에 의하여 국가가 변호인을 붙인다.', '⑤ 누구든지 체포 또는 구속의 이유와 변호인의 조력을 받을 권리가 있음을 고지받지 아니하고는 체포 또는 구속을 당하지 아니한다. 체포 또는 구속을 당한 자의 가족등 법률이 정하는 자에게는 그 이유와 일시·장소가 지체없이 통지되어야 한다.', '⑥ 누구든지 체포 또는 구속을 당한 때에는 적부의 심사를 법원에 청구할 권리를 가진다.', '⑦ 피고인의 자백이 고문·폭행·협박·구속의 부당한 장기화 또는 기망 기타의 방법에 의하여 자의로 진술된 것이 아니라고 인정될 때 또는 정식재판에 있어서 피고인의 자백이 그에게 불리한 유일한 증거일 때에는 이를 유죄의 증거로 삼거나 이를 이유로 처벌할 수 없다.', '② 모든 국민은 소급입법에 의하여 참정권의 제한을 받거나 재산권을 박탈당하지 아니한다.', '③ 모든 국민은 자기의 행위가 아닌 친족의 행위로 인하여 불이익한 처우를 받지 아니한다.', '② 국교는 인정되지 아니하며, 종교와 정치는 분리된다.', '② 언론·출판에 대한 허가나 검열과 집회·결사에 대한 허가는 인정되지 아니한다.']
```

In [105]:
```
# 전문+조+항 합치기
const_text_raw = preface_text + provision_li_text + provision_dd_text
print(const_text_raw[:20])
```

```
['유구한 역사와 전통에 빛나는 우리 대한국민은 3·1운동으로 건립된 대한민국임시정부의 법통과 불의에 항거한 4·19민주이념을 계승하고, 조국의 민주개혁과 평화적 통일의 사명에 입각하여 정의·인도와 동포애로써 민족의 단결을 공고히 하고, 모든 사회적 폐습과 불의를 타파하며, 자율과 조화를 바탕으로 자유민주적 기본질서를 더욱 확고히 하여 정치·경제·사회·문화의 모든 영역에 있어서 각인의 기회를 균등히 하고, 능력을 최고도로 발휘하게 하며, 자유와 권리에 따르는 책임과 의무를 완수하게 하여, 안으로는 국민생활의 균등한 향상을 기하고 밖으로는 항구적인 세계평화와 인류공영에 이바지함으로써 우리들과 우리들의 자손의 안전과 자유와 행복을 영원히 확보할 것을 다짐하면서 1948년 7월 12일에 제정되고 8차에 걸쳐 개정된 헌법을 이제 국회의 의결을 거쳐 국민투표에 의하여 개정한다.', '제1조 ① 대한민국은 민주공화국이다.', '제2조 ① 대한민국의 국민이 되는 요건은 법률로 정한다.', '제3조 대한민국의 영토는 한반도와 그 부속도서로 한다.', '제4조 대한민국은 통일을 지향하며, 자유민주적 기본질서에 입각한 평화적 통일 정책을 수립하고 이를 추진한다.', '제5조 ① 대한민국은 국제평화의 유지에 노력하고 침략적 전쟁을 부인한다.', '제6조 ① 헌법에 의하여 체결·공포된 조약과 일반적으로 승인된 국제법규는 국내법과 같은 효력을 가진다.', '제7조 ① 공무원은 국민전체에 대한 봉사자이며, 국민에 대하여 책임을 진다.', '제8조 ① 정당의 설립은 자유이며, 복수정당제는 보장된다.', '제9조 국가는 전통문화의 계승·발전과 민족문화의 창달에 노력하여야 한다.', '제10조 모든 국민은 인간으로서의 존엄과 가치를 가지며, 행복을 추구할 권리를 가진다. 국가는 개인이 가지는 불가침의 기본적 인권을 확인하고 이를 보장할 의무를 진다.', '제11조 ① 모든 국민은 법 앞에 평등하다. 누구든지 성별·종교 또는 사회적 신분에 의하여 정치적·경제적·사회적·문화적 생활의 모든 영역에 있어서 차별을 받지 아니한다.', '제12조 ① 모든 국민은 신체의 자유를 가진다. 누구든지 법률에 의하지 아니하고는 체포·구속·압수·수색 또는 심문을 받지 아니하며, 법률과 적법한 절차에 의하지 아니하고는 처벌·보안처분 또는 강제노역을 받지 아니한다.', '제13조 ① 모든 국민은 행위시의 법률에 의하여 범죄를 구성하지 아니하는 행위로 소추되지 아니하며, 동일한 범죄에 대하여 거듭 처벌받지 아니한다.', '제14조 모든 국민은 거주·이전의 자유를 가진다.', '제15조 모든 국민은 직업선택의 자유를 가진다.', '제16조 모든 국민은 주거의 자유를 침해받지 아니한다. 주거에 대한 압수나 수색을 할 때에는 검사의 신청에 의하여 법관이 발부한 영장을 제시하여야 한다.', '제17조 모든 국민은 사생활의 비밀과 자유를 침해받지 아니한다.', '제18조 모든 국민은 통신의 비밀을 침해받지 아니한다.', '제19조 모든 국민은 양심의 자유를 가진다.']
```

2. 데이터 전처리

- 우리가 요리를 하기전 재료를 손질하듯이 데이터를 분석하기 전에도 데이터를 다듬어야 하고, 이 과정을 데이터 전처리 또는 데이터 클렌징이라고 함.
- 전 단계에서 텍스트를 수집했지만 각종 불용어(조·항번호, 개행문자, 숫자 및 기호, 불필요한 공백 등)가 제거되지 않았기 때문에 정제가 필요

In [109]:
```
import pandas as pd

# 텍스트 리스트 series로 변환
const_text_raw = pd.Series(const_text_raw)
print(const_text_raw.head(20))
```

```
0     유구한 역사와 전통에 빛나는 우리 대한국민은 3·1운동으로 건립된 대한민국임시정부의...
1                                제1조 ① 대한민국은 민주공화국이다.
2                     제2조 ① 대한민국의 국민이 되는 요건은 법률로 정한다.
3                      제3조 대한민국의 영토는 한반도와 그 부속도서로 한다.
4     제4조 대한민국은 통일을 지향하며, 자유민주적 기본질서에 입각한 평화적 통일 정책을...
5              제5조 ① 대한민국은 국제평화의 유지에 노력하고 침략적 전쟁을 부인한다.
6     제6조 ① 헌법에 의하여 체결·공포된 조약과 일반적으로 승인된 국제법규는 국내법과 ...
7            제7조 ① 공무원은 국민전체에 대한 봉사자이며, 국민에 대하여 책임을 진다.
8                    제8조 ① 정당의 설립은 자유이며, 복수정당제는 보장된다.
9              제9조 국가는 전통문화의 계승·발전과 민족문화의 창달에 노력하여야 한다.
10    제10조 모든 국민은 인간으로서의 존엄과 가치를 가지며, 행복을 추구할 권리를 가진...
11    제11조 ① 모든 국민은 법 앞에 평등하다. 누구든지 성별·종교 또는 사회적 신분에...
12    제12조 ① 모든 국민은 신체의 자유를 가진다. 누구든지 법률에 의하지 아니하고는 ...
13    제13조 ① 모든 국민은 행위시의 법률에 의하여 범죄를 구성하지 아니하는 행위로 소...
14                         제14조 모든 국민은 거주·이전의 자유를 가진다.
15                          제15조 모든 국민은 직업선택의 자유를 가진다.
16    제16조 모든 국민은 주거의 자유를 침해받지 아니한다. 주거에 대한 압수나 수색을 ...
17                 제17조 모든 국민은 사생활의 비밀과 자유를 침해받지 아니한다.
18                      제18조 모든 국민은 통신의 비밀을 침해받지 아니한다.
19                           제19조 모든 국민은 양심의 자유를 가진다.
dtype: object
```

In [110]:
```
import re
import swifter

# 불용어(조항번호, 기호, 공백 등) 정규표현식 정의
pat_numbering = re.compile("제\Wd{1,}조")
pat_nonchar = re.compile("[^가-힣]")
pat_whitespaces = re.compile("\Ws{2,}")

# 불용어 제거
const_text_cleansed = (
     const_text_raw.swifter.apply(lambda x: re.sub(pat_numbering, " ", x))
    .swifter.apply(lambda x: re.sub(pat_nonchar, " ", x))
    .swifter.apply(lambda x: re.sub(pat_whitespaces, " ", x))
    .swifter.apply(lambda x: x.strip())
    )

print(const_text_cleansed.head(20))
```

```
0     유구한 역사와 전통에 빛나는 우리 대한국민은 운동으로 건립된 대한민국임시정부의 법통...
1                                  대한민국은 민주공화국이다
2                       대한민국의 국민이 되는 요건은 법률로 정한다
3                        대한민국의 영토는 한반도와 그 부속도서로 한다
4     대한민국은 통일을 지향하며 자유민주적 기본질서에 입각한 평화적 통일 정책을 수립하고...
5                 대한민국은 국제평화의 유지에 노력하고 침략적 전쟁을 부인한다
6     헌법에 의하여 체결 공포된 조약과 일반적으로 승인된 국제법규는 국내법과 같은 효력을...
7                  공무원은 국민전체에 대한 봉사자이며 국민에 대하여 책임을 진다
8                      정당의 설립은 자유이며 복수정당제는 보장된다
9                 국가는 전통문화의 계승 발전과 민족문화의 창달에 노력하여야 한다
```

```
10    모든 국민은 인간으로서의 존엄과 가치를 가지며 행복을 추구할 권리를 가진다 국가는 ...
11    모든 국민은 법 앞에 평등하다 누구든지 성별 종교 또는 사회적 신분에 의하여 정치적...
12    모든 국민은 신체의 자유를 가진다 누구든지 법률에 의하지 아니하고는 체포 구속 압수...
13    모든 국민은 행위시의 법률에 의하여 범죄를 구성하지 아니하는 행위로 소추되지 아니하...
14                                모든 국민은 거주 이전의 자유를 가진다
15                                 모든 국민은 직업선택의 자유를 가진다
16    모든 국민은 주거의 자유를 침해받지 아니한다 주거에 대한 압수나 수색을 할 때에는 ...
17                           모든 국민은 사생활의 비밀과 자유를 침해받지 아니한다
18                               모든 국민은 통신의 비밀을 침해받지 아니한다
19                                    모든 국민은 양심의 자유를 가진다
dtype: object
```

In [111]:
```
# 텍스트 합치기
text_pooled = " ".join(const_text_cleansed)
print(text_pooled[:500])
```

```
유구한 역사와 전통에 빛나는 우리 대한국민은 운동으로 건립된 대한민국임시정부의 법통과 불의에 항거한 민주이념을 계승하고 조국의 민주개혁과 평화적 통일의 사명에 입각하여 정의 인도와 동포애로써 민족의 단결을 공고히 하고 모든 사회적 폐습과 불의를 타파하며 자율과 조화를 바탕으로 자유민주적 기본질서를 더욱 확고히 하여 정치 경제 사회 문화의 모든 영역에 있어서 각인의 기회를 균등히 하고 능력을 최고도로 발휘하게 하며 자유와 권리에 따르는 책임과 의무를 완수하게 하여 안으로는 국민생활의 균등한 향상을 기하고 밖으로는 항구적인 세계평화와 인류공영에 이바지함으로써 우리들과 우리들의 자손의 안전과 자유와 행복을 영원히 확보할 것을 다짐하면서 년 월 일에 제정되고 차에 걸쳐 개정된 헌법을 이제 국회의 의결을 거쳐 국민투표에 의하여 개정한다 대한민국은 민주공화국이다 대한민국의 국민이 되는 요건은 법률로 정한다 대한민국의 영토는 한반도와 그 부속도서로 한다 대한민국은 통일을 지향하며 자유민주적 기본
```

In [112]:
```
# 토크나이징
from kiwipiepy import Kiwi
kiwi = Kiwi()
kiwi.prepare()
pos_tagging = kiwi.analyze(text_pooled)[0][0]
print(pos_tagging[:20])
```

```
[('유구', 'XR', 0, 2), ('하', 'XSA', 2, 1), ('ㄴ', 'ETM', 3, 0), ('역사', 'NNG', 4, 2), ('와', 'JC', 6, 1), ('전통', 'NNG', 8, 2), ('에', 'JKB', 10, 1), ('빛나', 'VV', 12, 2), ('는', 'ETM', 14, 1), ('우리', 'NP', 16, 2), ('대한', 'NNP', 19, 2), ('국민', 'NNG', 21, 2), ('은', 'JX', 23, 1), ('운동', 'NNG', 25, 2), ('으로', 'JKB', 27, 2), ('건립', 'NNG', 30, 2), ('되', 'XSV', 32, 1), ('ㄴ', 'ETM', 33, 0), ('대한민국', 'NNP', 34, 4), ('임시', 'NNG', 38, 2)]
```

In [113]:
```
# 명사추출
pat_noun = re.compile("NN[GP]")
nouns = []
for token, pos, _, _ in pos_tagging:
    if bool(re.match(pat_noun, pos)) and len(token) > 1:
        nouns.append(token)
print(nouns[:50])
```

```
['역사', '전통', '대한', '국민', '운동', '건립', '대한민국', '임시', '정부', '법통', '불의', '항거', '민주', '이념', '계승', '조국', '민주개혁', '평화', '통일', '사명', '입각', '정의', '인도', '동포애', '민족', '단결', '공고', '사회', '폐습', '불의', '타파', '자율', '조화', '바탕', '자유', '민주', '기본', '질서', '정치', '경제', '사회', '문화', '영역', '기회', '균등', '능력', '고도', '발휘', '자유', '권리']
```

3. 데이터 분석, 시각화 및 해석

1. 빈도분석

In [114]:
```
import numpy as np

# 기술통계량
print("총 단어 개수: "+str(len(nouns))+"개")
print("총 어휘 수(고유단어): "+str(len(np.unique(nouns)))+"개")
```

```
총 단어 개수: 3238개
총 어휘 수(고유단어): 803개
```

In [115]:
```
from collections import Counter

# 빈도수 확인
pd.set_option('display.max_rows', None)
count = Counter(nouns)
words = dict(count.most_common())
words_df = pd.DataFrame(word, index=["빈도"]).T
words_df.head(20)
```

Out[115]:

	빈도
법률	124
대통령	86
국가	76
헌법	73
국민	69
국회	69
필요	30
기타	30
선거	27
사항	26
법원	26
보장	24
의원	24
법관	23
회의	23
정부	22
임명	22
자유	21
임기	21
경제	20

2. 시각화

In [116]:
```
from wordcloud import WordCloud
import matplotlib.pyplot as plt

# 한글 폰트 설정
from matplotlib import font_manager
import matplotlib
import os

font_relpath = os.path.join("src", "font.ttf")
font_abspath = os.path.join(os.getcwd(), font_relpath)
```

In [117]:
```
# 워드클라우드 객체 생성
wordcloud = WordCloud(
    font_path = font_abspath,
    background_color="white",
    colormap= "Accent_r",
```

```
    width = 1500,
    height = 1500)

wordcloud_words = wordcloud.generate_from_frequencies(words)
wordcloud_array = wordcloud.to_array()
```

In [118]:

```
# 워드클라우드 그리기
fig = plt.figure(figsize=(20, 20))
plt.imshow(wordcloud_array)
plt.axis("off")
plt.show()
fig.savefig("대한민국 헌법.png")
```

3. 해석

- 데이터 과학 프로세스의 마지막은 분석 결과를 해석하는 것
- 이 과정에서 해당 분야(domain)에 대한 전문지식이 중요

> 분석 결과에 대한 법조인 선생님 여러분들의 생각은 어떠신가요?