

# 영화 추천 시스템과 데이터 수집 개요

개인화된 영화 추천 시스템은 현대 엔터테인먼트 플랫폼의 핵심 요소입니다. 사용자의 취향을 분석하고 최적화된 콘텐츠를 제안하는 이 시스템은 시청 경험을 혁신적으로 개선합니다.

이 프레젠테이션에서는 영화 추천 시스템의 핵심 구성 요소인 알고리즘 설계와 데이터 수집 방법에 대해 상세히 살펴보겠습니다.

# 1. 영화 추천 개요

영화 추천 시스템은 사용자의 취향과 시청 기록을 분석하여 개인화된 영화를 추천하는 지능형 시스템입니다. 넷플릭스, 왓챗과 같은 스트리밍 서비스에서 핵심적인 역할을 담당하며, 사용자의 만족도를 높이고 콘텐츠 소비를 촉진합니다.

## 영화 추천 시스템의 개요와 활용 사례

### 영화 추천 시스템이란?

영화 추천 시스템은 개인의 취향에 맞는 영화를 추천해주는 기술입니다.

넷플릭스에서 "당신을 위한 추천"이라는 문구를 본 적 있거나, 유튜브에서 "다음에 볼만한 영상"을 제안받은 경험이 있다면, 이것이 바로 추천 시스템의 결과입니다.

### 영화 추천 시스템의 필요성

- 1. 선택의 어려움 해결:** 세상에는 너무 많은 영화가 있습니다. 추천 시스템은 그중에서 사용자가 좋아할 만한 영화를 빠르게 찾아줍니다.
- 2. 시간 절약:** 사용자가 영화를 고르는 데 드는 시간을 줄여줍니다.
- 3. 맞춤형 경험 제공:** 사용자의 취향을 분석해 딱 맞는 영화를 추천함으로써 만족도를 높입니다.

## 영화 추천 시스템의 실제 활용 사례

### 넷플릭스

- **사용 방법:** 넷플릭스는 사용자가 시청한 영화와 좋아하는 영화 데이터를 분석하여 비슷한 영화를 추천합니다.
- 예를 들어, 액션 영화를 자주 본다면 다른 인기 액션 영화를 추천합니다.
- **효과:** 추천 시스템 덕분에 사용자는 넷플릭스에서 더 오랜 시간을 보내며 만족도가 높아집니다.

### 왓차

- **사용 방법:** 왓차는 사용자가 남긴 별점 데이터를 기반으로 영화 추천을 진행합니다.
- 예를 들어, "겨울왕국"에 높은 별점을 줬다면, 비슷한 애니메이션 영화를 추천합니다.
- **효과:** 사용자는 새로운 영화를 쉽게 발견할 수 있으며, 플랫폼의 이용이 증가합니다.

### 유튜브

- **사용 방법:** 유튜브는 사용자의 시청 기록을 분석하여 비슷한 주제의 동영상을 추천합니다.
- 예를 들어, "영화 리뷰" 영상을 여러 번 본 경우, 다른 영화 리뷰 영상을 추천합니다.
- **효과:** 사용자는 유튜브에 더 오래 머무르며 만족감이 증가합니다.

## 2. 영화 추천 알고리즘

영화 추천 알고리즘은 사용자의 취향과 패턴을 분석하여 최적의 영화를 제안하는 기술적 방법입니다. 주로 콘텐츠 기반 추천과 협업 필터링, 두 가지 주요 방식으로 구현됩니다. 이러한 알고리즘을 통해 사용자는 자신의 취향에 맞는 새로운 영화를 쉽게 발견할 수 있습니다.

## 영화 추천 알고리즘이란?

영화 추천 알고리즘이란 **사용자가 좋아할 만한 영화를 예측하고 추천하는 기술**을 의미합니다.

친구에게 "재미있는 영화 추천해줘!"라고 부탁하는 것처럼, 컴퓨터가 데이터를 분석해 영화를 추천해주는 것입니다.

이러한 알고리즘은 주로 **콘텐츠 기반 추천**과 **협업 필터링**이라는 두 가지 방법으로 구현됩니다.

# 콘텐츠 기반 추천 알고리즘

## 콘텐츠 기반 추천이란?

- **추천 기준:** 영화의 **내용(특징)**을 활용하여 비슷한 영화를 추천합니다.
- **작동 원리:** 사용자가 좋아한 영화의 장르, 감독, 배우, 줄거리 등 특징을 분석한 뒤, 이와 유사한 특징을 가진 다른 영화를 추천합니다.

## 예시

- 사용자가 "겨울왕국"을 좋아한다고 입력했을 경우:
  - 같은 장르(애니메이션, 가족)나 같은 제작사(디즈니)의 "라푼젤"을 추천.
  - 줄거리에서 비슷한 특징(모험, 가족애)을 가진 "인어공주"를 추천.

## 비슷한 줄거리를 가진 영화 추천의 예

- 사용자가 "겨울왕국"을 좋아한다면, "겨울왕국"처럼 자매 간의 우애를 그린 영화나 모험을 떠나는 이야기를 가진 영화를 추천할 수 있습니다.
  - 예: "코코" (가족의 중요성을 다룸), "업" (모험과 감동).

## 콘텐츠 기반 추천의 장점과 단점

- **장점:**
  - 사용자가 좋아한 영화의 특징을 분석하기 때문에 개인화가 잘 이루어집니다.
  - 다른 사람의 데이터에 의존하지 않아도 추천이 가능합니다.
- **단점:**
  - 사용자가 본 영화와 비슷한 영화만 추천하므로, 새로운 장르를 발견하기 어렵습니다.

# 협업 필터링 추천 알고리즘

## 협업 필터링이란?

- **추천 기준:** 다른 사용자와의 **취향 유사성**을 바탕으로 영화를 추천합니다.
- **작동 원리:** 비슷한 영화를 좋아한 사용자들의 데이터를 비교하여 추천합니다.

## 예시

- 사용자 A가 "겨울왕국"과 "라푼젤"을 좋아하고, 사용자 B가 "겨울왕국"과 "업"을 좋아한다고 가정합니다.
- 사용자가 "겨울왕국"을 좋아한다고 입력하면:
  - 사용자 B가 좋아한 "업"을 추천합니다.

## 협업 필터링의 장점과 단점

- **장점:**
  - 영화의 내용과 상관없이 다양한 영화를 추천할 수 있습니다.
  - 새로운 장르나 예상치 못한 영화를 추천받을 가능성이 높습니다.
- **단점:**
  - 충분한 사용자 데이터가 없을 경우 추천이 어렵습니다
  - 사용자 데이터를 많이 수집해야 하므로 구현이 다소 복잡할 수 있습니다.



# 콘텐츠 기반 추천과 협업 필터링의 비교

| 구분       | 콘텐츠 기반 추천                          | 협업 필터링                      |
|----------|------------------------------------|-----------------------------|
| 추천 기준    | 영화 자체의 정보(장르, 배우, 감독, 줄거리 등)       | 다른 사용자와의 취향 유사성             |
| 장점       | 본 영화와 유사한 영화 추천에 적합                | 새로운 장르나 다양한 영화를 추천받을 가능성 높음 |
| 단점       | 새로운 장르나 예상치 못한 영화 추천이 어려움          | 사용자 데이터가 충분하지 않을 경우 추천이 어려움 |
| 추천 방식 예시 | "겨울왕국" → "라푼젤", "인어공주"(장르와 줄거리 유사) | "겨울왕국" 좋아하는 사용자가 추천한 "업" 추천 |

### 3. 영화 데이터 수집의 세계

영화 추천 시스템을 만들기 위해서는 먼저 방대한 양의 영화 데이터를 수집해야 합니다. 이를 위해 웹 스크래핑 도구인 BeautifulSoup과 Selenium을 활용하여 효율적으로 데이터를 수집할 수 있습니다. 이러한 도구들을 통해 영화 정보, 사용자 리뷰, 평점 등 다양한 데이터를 자동으로 수집하고 분석할 수 있습니다.

## 데이터 수집이란?

데이터 수집은 인터넷에서 정보를 가져와 저장하는 작업을 의미합니다.

예를 들어, 네이버 영화 사이트에서 영화 제목, 평점, 감독 정보를 가져오는 것입니다.

사람이 직접 일일이 웹사이트를 보고 정보를 복사해서 저장할 수도 있지만, **컴퓨터가 자동으로 정보를 가져오도록 프로그래밍**할 수 있습니다.

이 작업에 사용하는 도구 중 대표적인 것이 **BeautifulSoup**과 **Selenium**입니다.

# BeautifulSoup: 빠르고 간단한 데이터 수집 도구

## BeautifulSoup은 무엇인가요?

- BeautifulSoup은 웹사이트의 HTML(웹 페이지를 구성하는 코드)을 읽고, 필요한 정보를 쉽게 찾을 수 있도록 도와주는 도구입니다.
- **비유:** 웹사이트가 책이라면, BeautifulSoup은 목차를 보고 필요한 내용을 빨리 찾아주는 도구입니다.

## BeautifulSoup이 적합한 경우

- 웹사이트 구조가 단순하고, JavaScript 같은 복잡한 동작 없이 데이터를 바로 볼 수 있을 때 사용합니다.

## 어떻게 사용하나요?

1. 컴퓨터가 웹사이트의 내용을 가져옵니다. (이를 **HTML 파일**이라고 부릅니다.)
2. BeautifulSoup이 HTML 파일에서 필요한 부분(예: 영화 제목, 평점)을 찾아줍니다.
3. 찾아낸 데이터를 저장하거나 활용합니다.

# Selenium: 복잡한 웹사이트 데이터 수집 도구

## Selenium은 무엇인가요?

- Selenium은 사람이 웹사이트를 클릭하고 스크롤하는 동작을 대신해주는 도구입니다.
- **비유:** Selenium은 컴퓨터가 마치 사람처럼 웹사이트를 보고, 버튼을 누르고, 페이지를 넘기게 만드는 로봇과 같습니다.

## Selenium이 적합한 경우

- 웹사이트가 JavaScript를 사용해서 동적으로 데이터를 로드할 때.
- 로그인, 스크롤, 버튼 클릭 등 사람이 직접 해야 할 작업이 필요한 경우.

## 어떻게 사용하나요?

1. Selenium은 웹 브라우저(예: 크롬, 파이어폭스)를 자동으로 실행합니다.
2. 지정된 웹사이트에 접속하여 사람이 하는 것처럼 클릭하거나 정보를 확인합니다.
3. 필요한 데이터를 찾아 저장합니다.

# BeautifulSoup과 Selenium의 차이점

| 도구            | 사용 상황  | 장점                 | 단점                |
|---------------|--|--------------------|-------------------|
| BeautifulSoup | 단순한 HTML 구조에서 정보를 수집할 때                      | 빠르고 간단함            | 동적 웹사이트에서는 한계가 있음 |
| Selenium      | 버튼 클릭, 스크롤 등 동적 작업이 필요한 복잡한 웹사이트에서 정보를 수집할 때 | 사람처럼 웹사이트를 다룰 수 있음 | 느리고, 리소스 소모가 큼    |

## BeautifulSoup과 Selenium을 함께 사용하는 이유

때로는 단순한 작업은 BeautifulSoup으로 처리하고, 복잡한 작업은 Selenium으로 처리해야 할 때가 있습니다.

예를 들어:

1. Selenium으로 네이버 영화에 로그인하고, 특정 영화의 상세 페이지로 이동합니다.
2. BeautifulSoup으로 영화 제목, 평점 등의 정보를 가져옵니다.

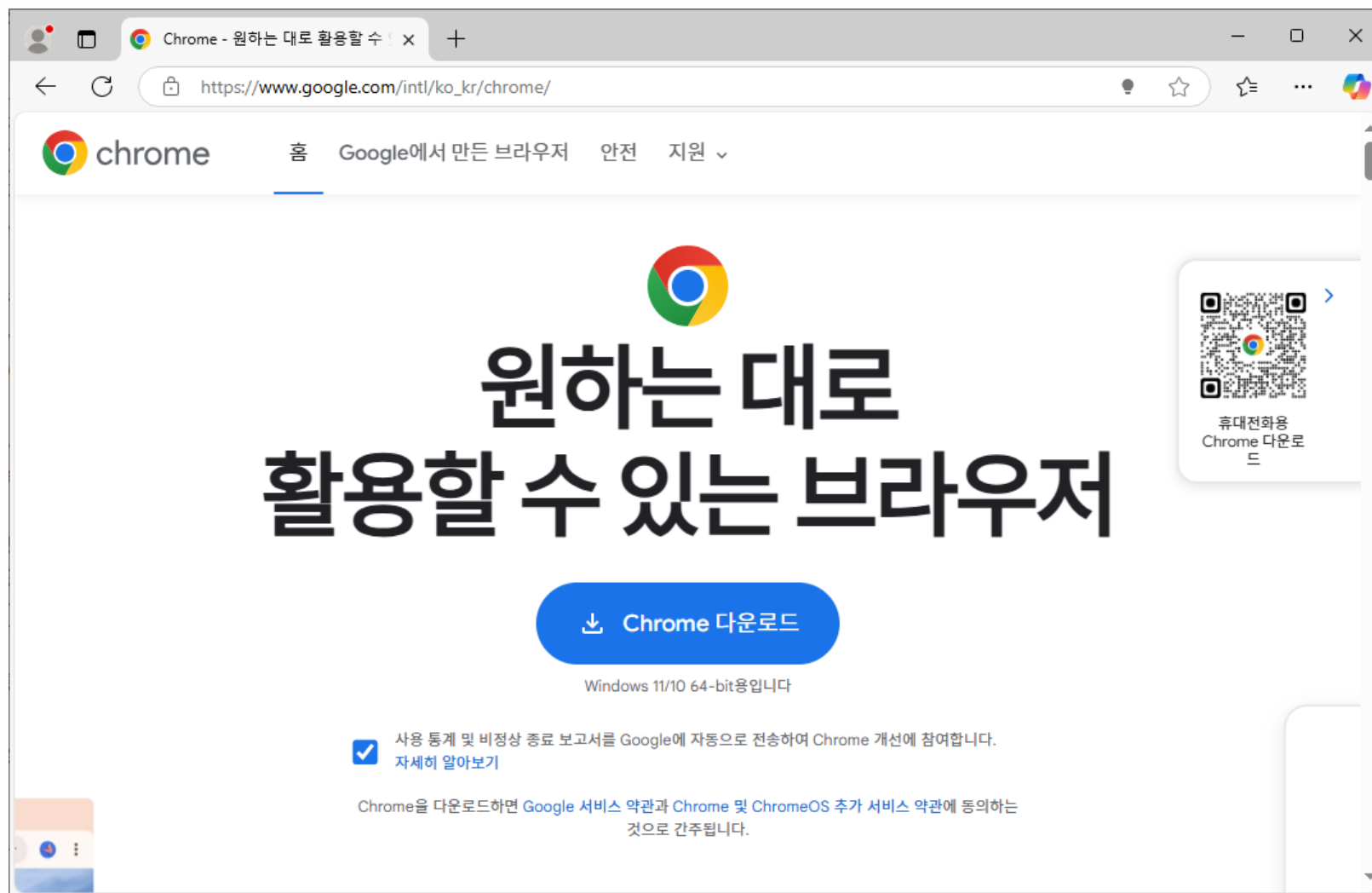
## 4. 크롬 설치

Selenium 을 이용해서 데이터를 다운로드 받기 위해서 크롬을 설치



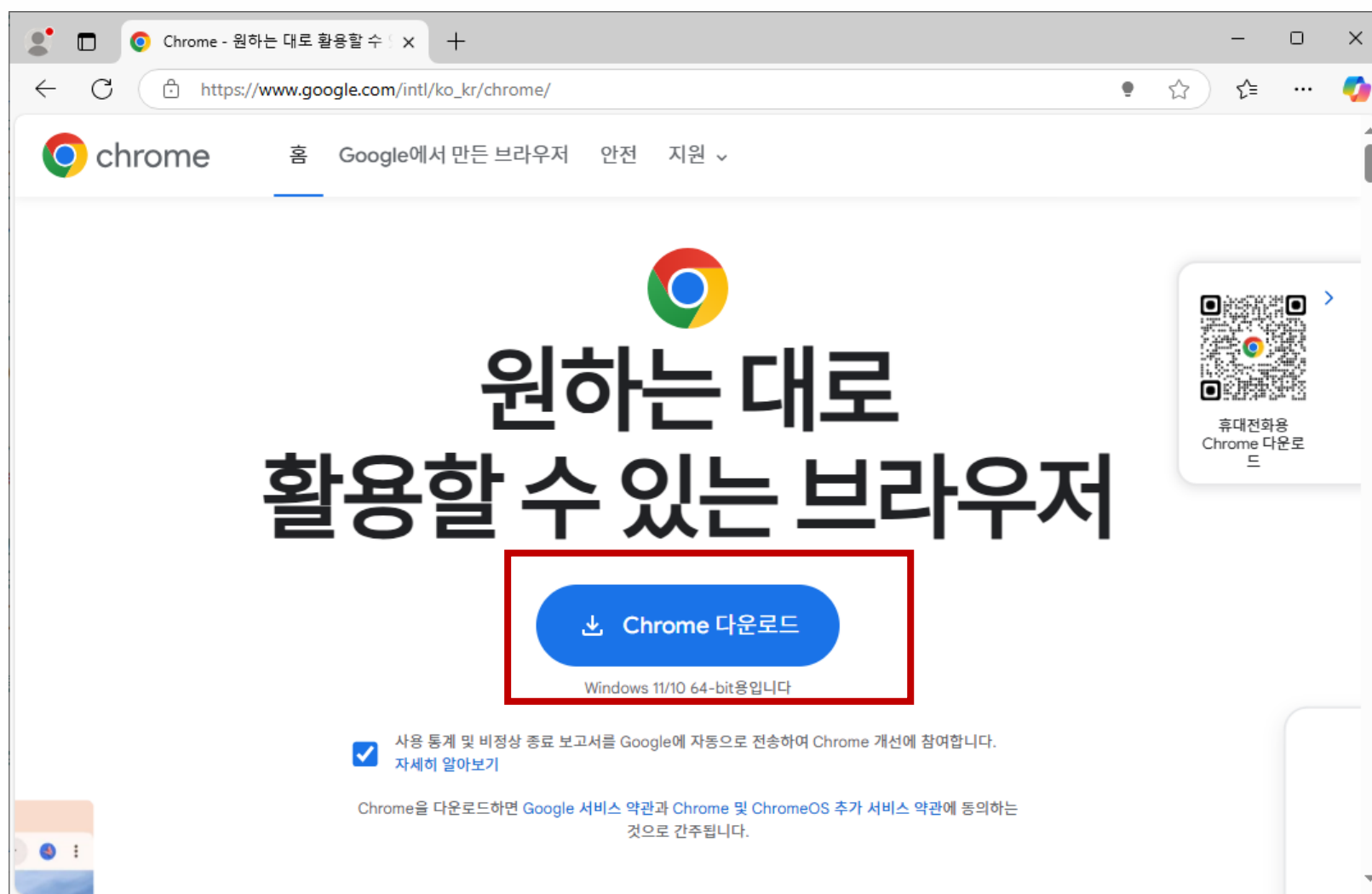
# 크롬 설치

[https://www.google.com/intl/ko\\_kr/chrome/](https://www.google.com/intl/ko_kr/chrome/) 접속 합니다

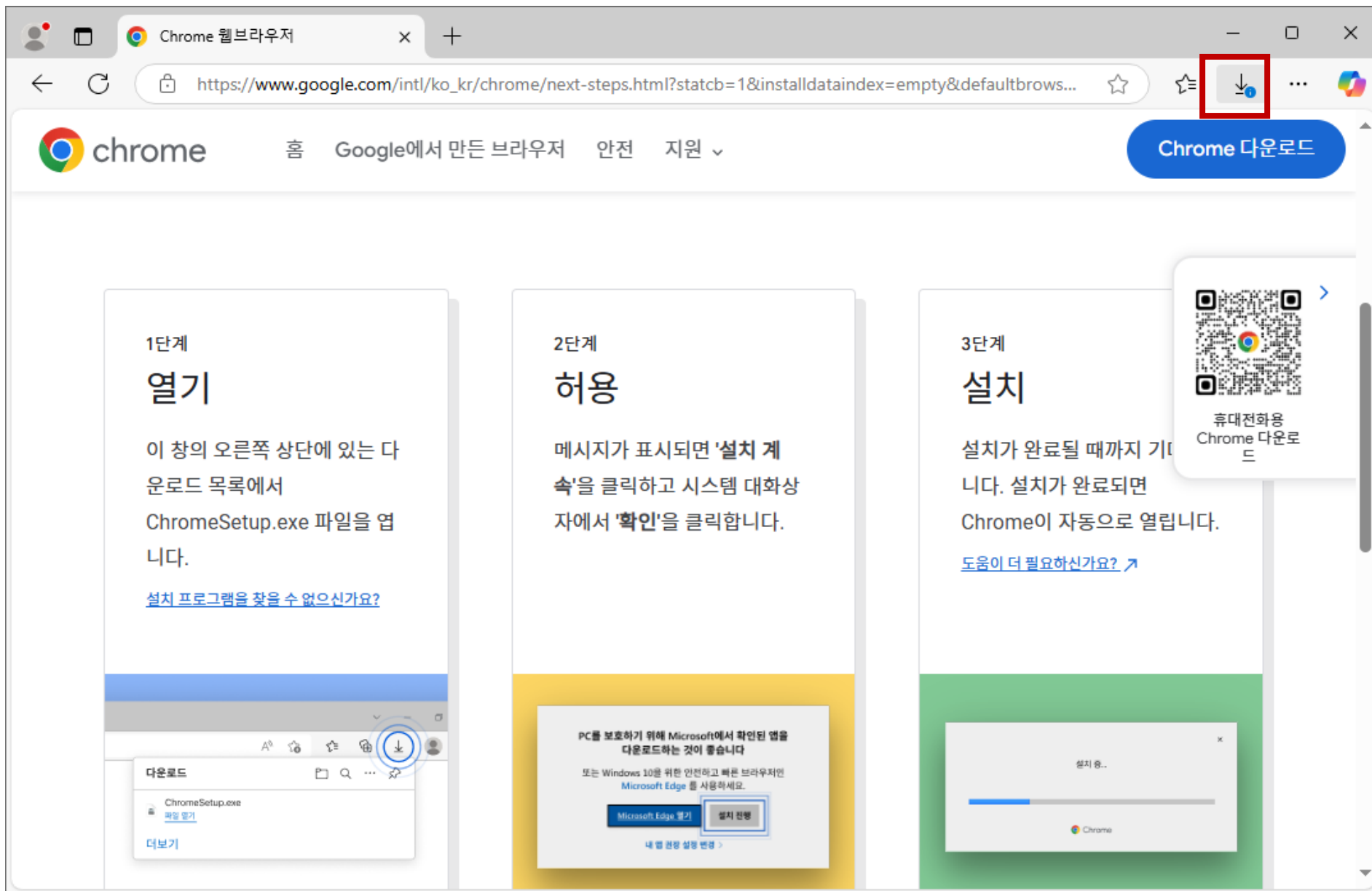


# 크롬 설치

크롬 다운로드를 선택 합니다

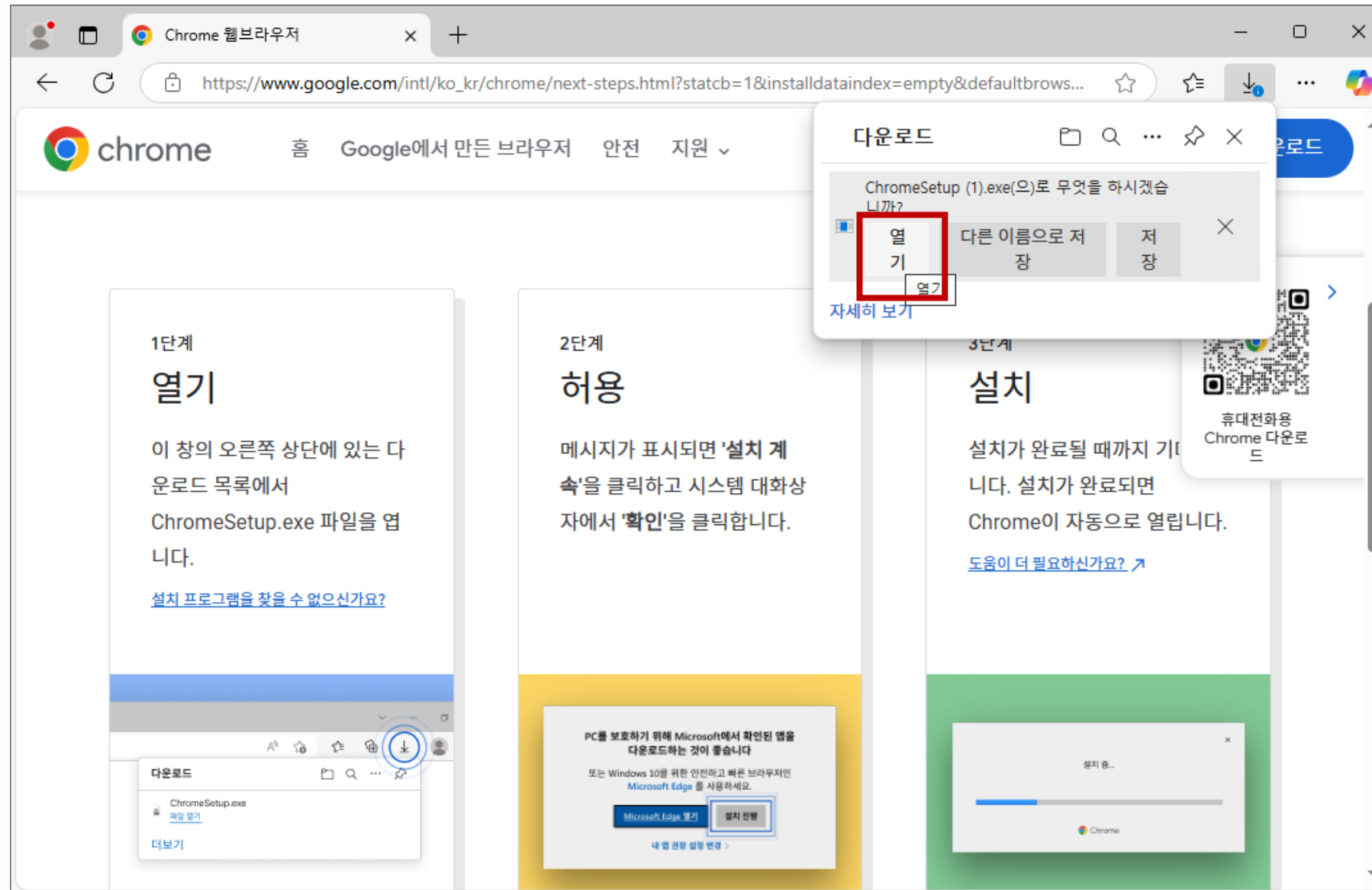


# 크롬 설치



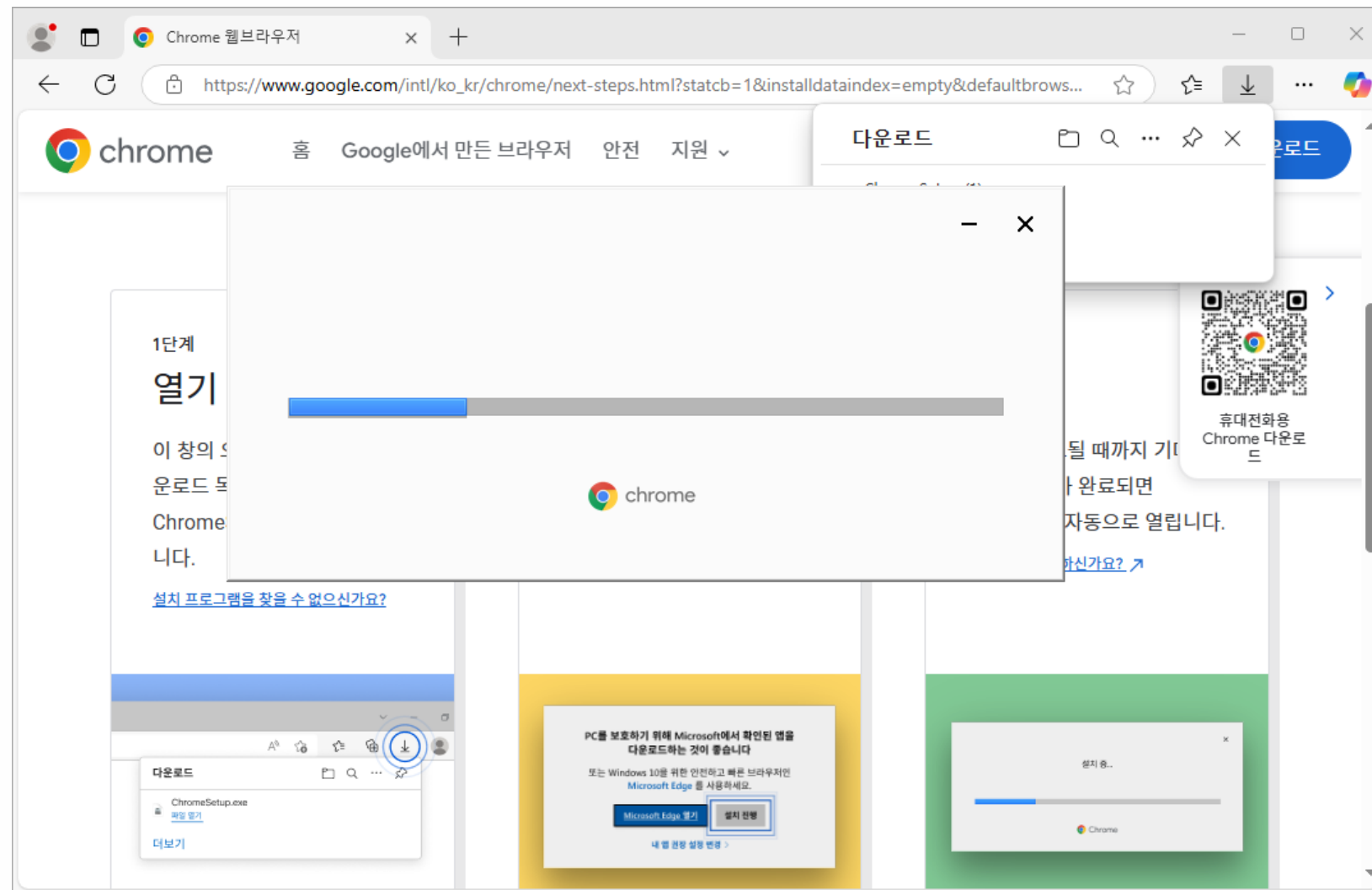
# 크롬 설치

열기를 선택 합니다



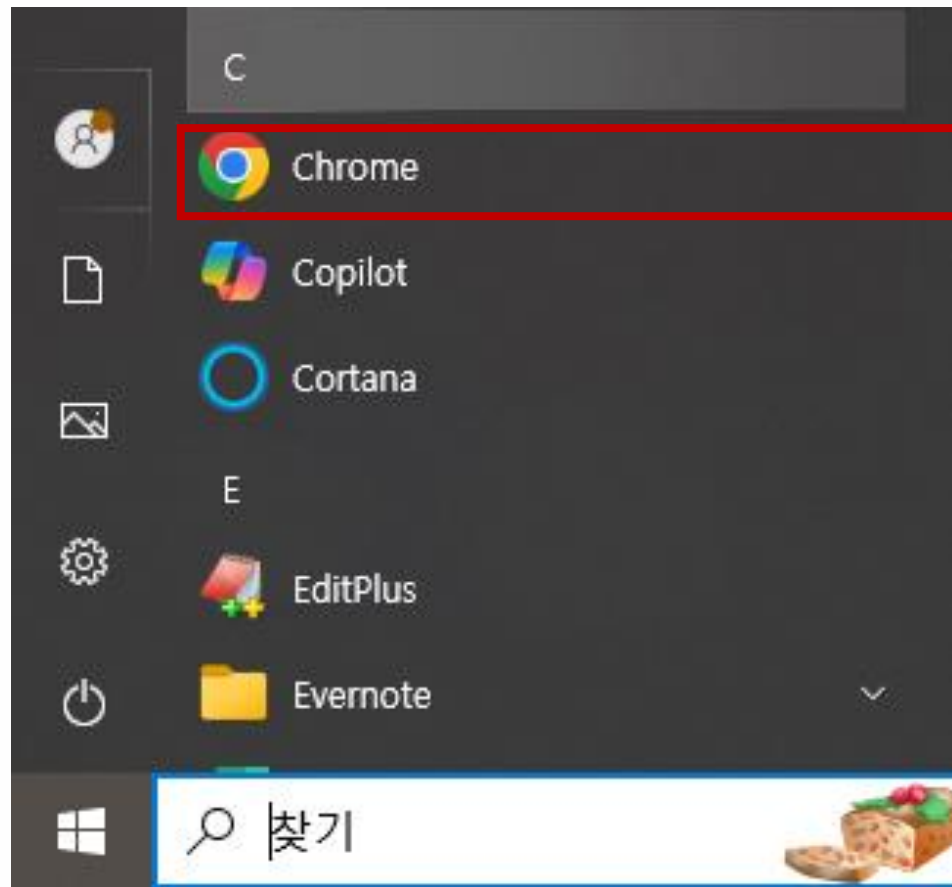
# 크롬 설치

설치를 진행 합니다



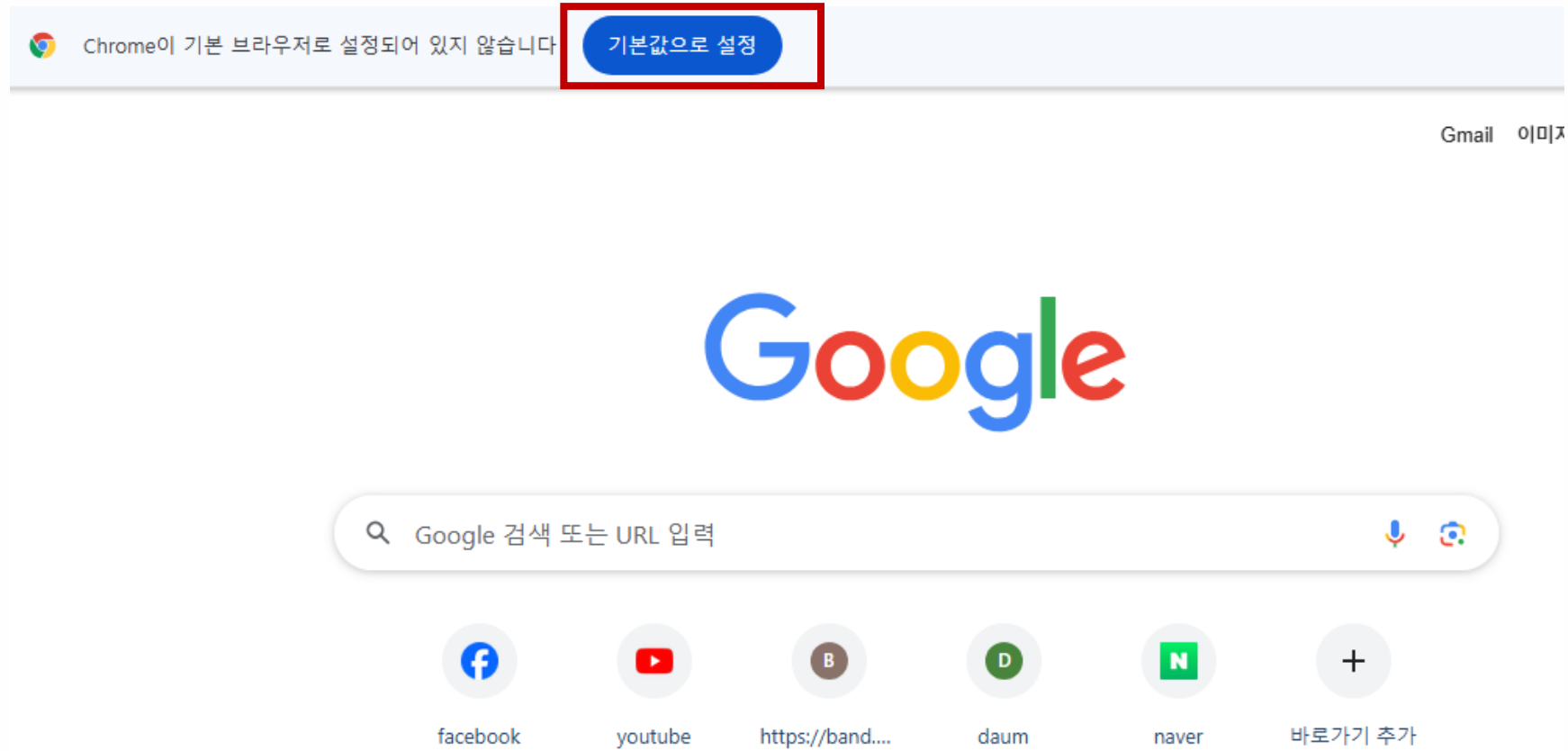
# 크롬 설치

크롬을 실행 합니다



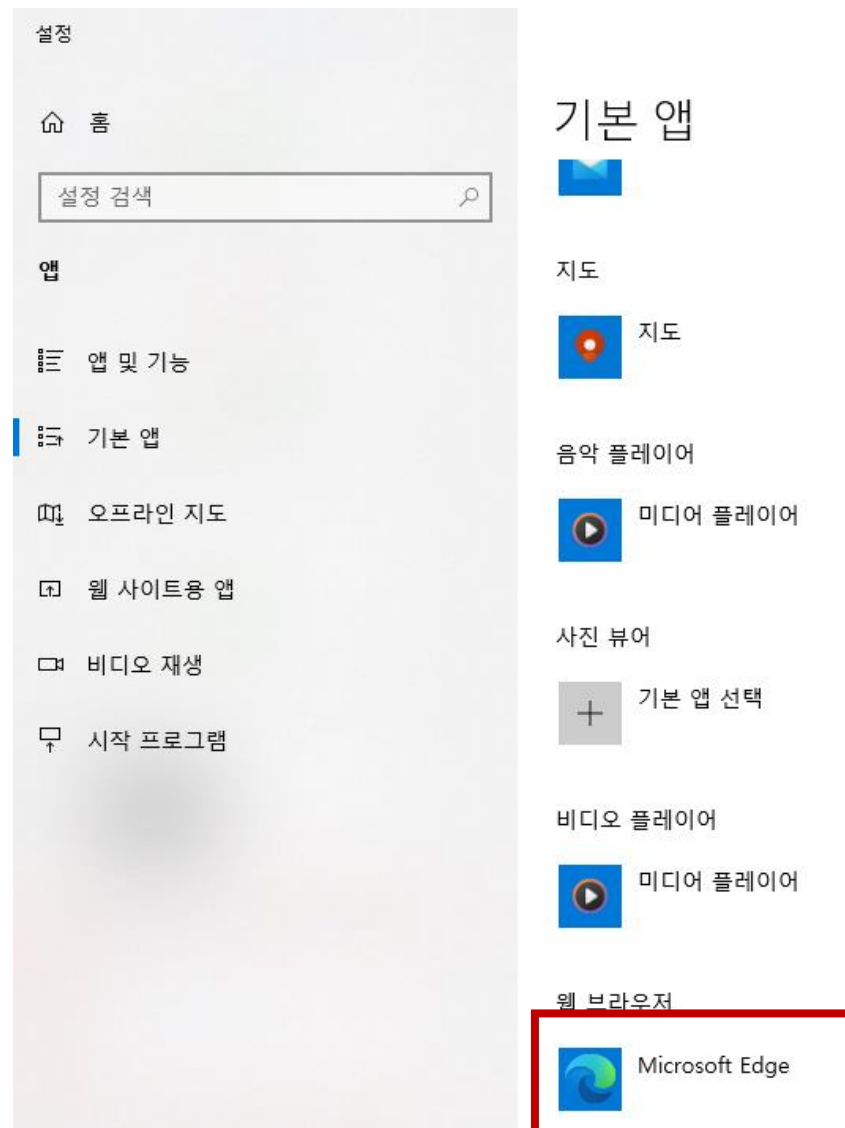
# 크롬 설치

기본 값으로 설정을 선택합니다



# 크롬 설치

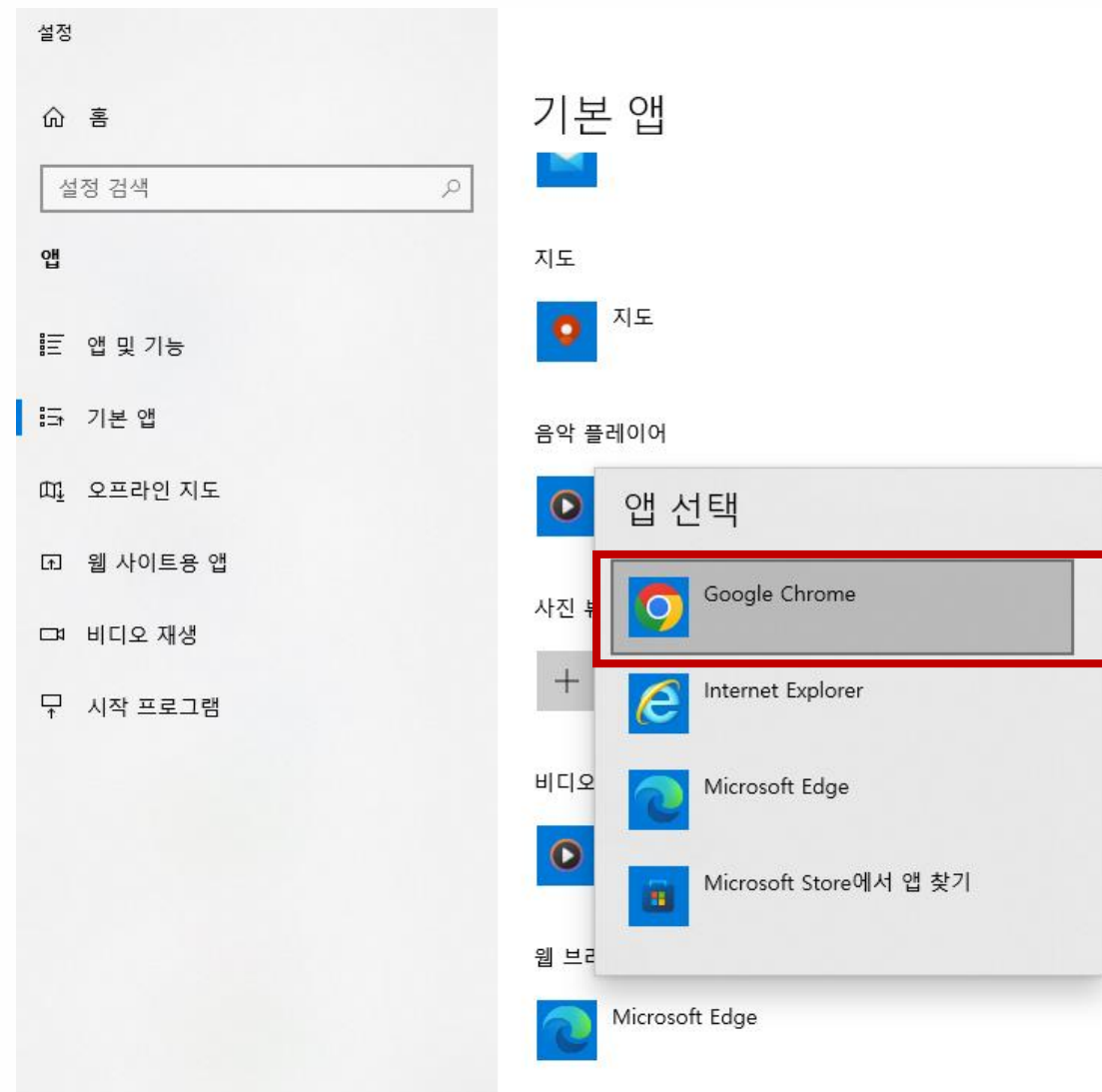
웹 브라우저를 선택 합니다





# 크롬 설치

크롬을 선택 합니다



# 크롬 설치

크롬으로 변경 되었는지 확인 합니다

