

Corpus:

[1]https://en.wikipedia.org/wiki/United_States

The United States of America (USA), commonly known as the United States (U.S. or US) or America, is a country comprising 50 states, a federal district, five major self-governing territories, and various possessions.[i] At 3.8 million square miles (9.8 million km²), the United States is the world's third or fourth largest country by total area[d] and is slightly smaller than the entire continent of Europe. With a population of over 327 million people, the U.S. is the third most populous country. The capital is Washington, D.C., and the most populous city is New York City. Most of the country is located contiguously in North America between Canada and Mexico.

[2]https://en.wikipedia.org/wiki/San_Jose,_California

San Jose[A] (/ˌsæn hoʊˈzeɪ, -ˈseɪ/; Spanish: [saŋ xoˈse]; Spanish for "Saint Joseph"),[13] officially the City of San José,[B] is the economic, cultural and political center of Silicon Valley, and the largest city in Northern California (both by population and area). With an estimated 2017 population of 1,035,317, it is the third-most populous city in California (after Los Angeles and San Diego) and the tenth-most populous in United States.[14] Located in the center of the Santa Clara Valley, on the southern shore of San Francisco Bay, San Jose covers an area of 179.97 square miles (466.1 km²). San Jose is the county seat of Santa Clara County, the most affluent county in California and one of the most affluent counties in the United States.[15][16][17][18] San Jose is the main component of the San Jose–Sunnyvale–Santa Clara Metropolitan Statistical Area, with an estimated population of around 2 million residents in 2018.[19] It is also the most populous city in both the San Francisco Bay Area and the San Jose-San Francisco-Oakland Combined Statistical Area, which contain 7.7 million and 8.7 million people respectively.

[3]https://en.wikipedia.org/wiki/Apple_Park

Apple Park is the corporate headquarters of Apple Inc., located at 1 Apple Park Way in Cupertino, California, United States. It was opened to employees in April 2017, while construction was still underway. Its research and development facilities are occupied by more than 2,000 people. It superseded the original headquarters at 1 Infinite Loop, which opened in 1993.

[4]https://en.wikipedia.org/wiki/Sequoia_National_Park

Sequoia National Park is an American national park in the southern Sierra Nevada east of Visalia, California. The park was established on September 25, 1890 to protect 404,064 acres (631 sq mi; 163,519 ha; 1,635 km²)[1] of forested mountainous terrain. Encompassing a vertical relief of nearly 13,000 feet (4,000 m), the park contains the highest point in the contiguous United States, Mount Whitney, at 14,505 feet (4,421 m) above sea level. The park is south of, and contiguous with, Kings Canyon National Park; the two parks are administered by the National Park Service together as the Sequoia and Kings Canyon National Parks. UNESCO designated the areas as Sequoia-Kings Canyon Biosphere Reserve in 1976.[3]

[5]<https://en.wikipedia.org/wiki/Water>

Water is a transparent, tasteless, odorless, and nearly colorless chemical substance, which is the main constituent of Earth's streams, lakes, and oceans, and the fluids of most living organisms. It is vital for all known forms of life, even though it provides no calories or organic nutrients. Its chemical formula is H₂O, meaning that each of its molecules contains one oxygen and two hydrogen atoms, connected by covalent bonds. Water is the name of the liquid state of H₂O at standard ambient temperature and pressure. It forms precipitation in the form of rain and aerosols in the form of fog. Clouds are formed from suspended

droplets of water and ice, its solid state. When finely divided, crystalline ice may precipitate in the form of snow. The gaseous state of water is steam or water vapor. Water moves continually through the water cycle of evaporation, transpiration (evapotranspiration), condensation, precipitation, and runoff, usually reaching the sea.

[6]<https://en.wikipedia.org/wiki/Photograph>

A photograph (also known as a photo) is an image created by light falling on a photosensitive surface, usually photographic film or an electronic image sensor, such as a CCD or a CMOS chip. Most photographs are created using a camera, which uses a lens to focus the scene's visible wavelengths of light into a reproduction of what the human eye would see. The process and practice of creating such images is called photography. The word photograph was coined in 1839 by Sir John Herschel and is based on the Greek φῶς (phos), meaning "light," and γραφή (graphê), meaning "drawing, writing," together meaning "drawing with light." [1]

[7]<https://en.wikipedia.org/wiki/Violin>

The violin, sometimes known as a fiddle, is a wooden string instrument in the violin family. Most violins have a hollow wooden body. It is the smallest and highest-pitched instrument in the family in regular use. Smaller violin-type instruments exist, including the violino piccolo and the kit violin, but these are virtually unused. The violin typically has four strings, usually tuned in perfect fifths with notes G3, D4, A4, E5, and is most commonly played by drawing a bow across its strings, though it can also be played by plucking the strings with the fingers (pizzicato) and by striking the strings with the wooden side of the bow (col legno).

[8]<https://en.wikipedia.org/wiki/Green>

Green is the color between blue and yellow on the visible spectrum. It is evoked by light which has a dominant wavelength of roughly 495–570 nm. In subtractive color systems, used in painting and color printing, it is created by a combination of yellow and blue, or yellow and cyan; in the RGB color model, used on television and computer screens, it is one of the additive primary colors, along with red and blue, which are mixed in different combinations to create all other colors. By far the largest contributor to green in nature is chlorophyll, the chemical by which plants photosynthesize and convert sunlight into chemical energy.

[9]https://en.wikipedia.org/wiki/Yosemite_National_Park

Yosemite National Park (/joʊˈsɛmɪti/ yoh-SEM-i-tee)[4] is an American national park located in the western Sierra Nevada of Central California y,[5][6] bounded on the southeast by Sierra National Forest and on the northwest by Stanislaus National Forest. The park is managed by the National Park Service and covers an area of 748,436 acres (1,169 sq mi; 3,029 km2)[2] and sits in four counties: centered in Tuolumne and Mariposa, extending north and east to Mono and south to Madera County. Designated a World Heritage site in 1984, Yosemite is internationally recognized for its granite cliffs, waterfalls, clear streams, giant sequoia groves, lakes, mountains, meadows, glaciers, and biological diversity.[7] Almost 95% of the park is designated wilderness.[8]

[10]<https://en.wikipedia.org/wiki/Headphones>

Headphones (or head-phones in the early days of telephony and radio) traditionally refer to a pair of small loudspeaker drivers worn on or around the head over a user's ears. They are electroacoustic transducers, which convert an electrical signal to a corresponding sound. Headphones let a single user listen to an audio source privately, in contrast to a loudspeaker, which emits sound into the open air for anyone nearby to hear. Headphones are also known as earspeakers, earphones[1] or, colloquially, cans.

Experiment Results:

Aim: Calculating recall@5 and precision@5

Total documents in corpus: 10

Sr. No.	Query	Rel	Res	Recall	Precision
		Relevant documents (by hand)	Total number of results returned by program	$\frac{ Rel }{ Res }$	$\frac{ Rel }{ Res }$
1.	_AND _AND National Park California	2 Doc id <4,9>	2 Doc id <4,9>	$2 / 2 = 1$	$2 / 2 = 1$
2.	_OR _OR Yosemite Park California	1 Doc id <9>	4 Doc id <2,3,4,9>	$1 / 1 = 1$	$1 / 4 = 0.25$

```
C:\Users\Harshit\Desktop\HW2>java PositiveRank HW2.txt 5 "_AND _AND National Park California"
Doc Id    Score
9         0.3304327651571065
4         0.32231618929831807

C:\Users\Harshit\Desktop\HW2>java PositiveRank HW2.txt 5 "_OR _OR Yosemite Park California"
Doc Id    Score
9         0.30206033260492543
4         0.10981109533487207
3         0.08681413504175049
2         0.02829625326157273
```

For Query 1: _AND _AND National Park California

Program returned 2 documents : 4,9
which has all the terms from the query.

Now, out of these doc

Doc id = 4 is about "Sequoia_National_Park" and
Doc id = 9 is about "Yosemite_National_Park"

Since, we wanted to search about National Parks so both of these are relevant to us,

So, $|Rel|=2$, $|Res|=2$

Hence, Recall = $2/2 = 1$ & Precision = $2/2 = 1$.

For Query 2: _OR _OR Yosemite Park California

Program returned 4 documents : 2,3,4,9
which has at least one term from the query.

Now, out of these doc

Doc id = 4 is about "Sequoia_National_Park" and
Doc id = 9 is about "Yosemite_National_Park"

But only Doc 9 (Yosemite) is relevant according to us, which is what we were looking for.

So, $|\text{Rel}|=1$, $|\text{Res}|=4$

Hence, Recall = $1/1 = 1$ & Precision = $1/4 = 0.25$

For Query 3: _OR _OR San Jose United States

```
C:\Users\Harshit\Desktop\HW2>java PositiveRank HW2.txt 5 "_OR _OR San Jose United States"
Doc Id   Score
2        0.38314927741888444
1        0.0921232726866131
3        0.04078663577425969
4        0.03033440976481485
```

Program returned 4 documents : 1,2,3,4
which has at least one term from the query.

Now, out of these doc

Doc id = 1 is about "United States" and
Doc id = 2 is about "San Jose"
Doc id = 3 is about "Apple Park" and
Doc id = 4 is about "Sequoia National Park"

Here, only Doc 2 (San Jose , US) is relevant according to us, which is what we were looking for.

So, $|\text{Rel}|=1$, $|\text{Res}|=4$

Hence, Recall = $1/1 = 1$ & Precision = $1/4 = 0.25$

Final Results:

Out of all the queries executed, conjunctive query performed the best with precision = 100%.