

LARGE-SCALE SEARCH TECHNOLOGIES

Why is building an internet search engine so hard?

Over 150 trillion webpages. Only 1 of every 3,000 webpages is indexed.

It is practically impossible to index all the words in every webpage

One cannot build a competitive search without huge amount of users' data



It's too expensive to compute search results on the fly



State of the art technologies wouldn't work to power large-scale web search engines





Scalability challenges



The core algorithm of an inverted index scheme is based on a list intersection

This is a computationally inefficient operation, which is impractical when it comes to the web and its billions of webpages.



Noise reduction challenges



Textual measures: Word intersections, word distances, word positions, td-idf, BM25 etc.



No matter which textually-based approach you take, it might work well on one query, but poorly on the other.



There is no point trying to index the parts of a webpage that are unlikely to appear in prospective users' search queries



Suppose that for each webpage you had a list of all the search queries that ended with a user clicking on this webpage (say from Google)

Any part of a webpage that does interact with any of the search queries leading to it is completely irrelevant for the retrieval process

 \bigcirc

Hence, you could focus on only indexing the search queries.



Example – Cyberpunk 2077







User queries

User Queries

Cyberpunk 2077 vehicles

Cyberpunk 2077 cars

Cyberpunk cars

Cyberpunk vehicles

Cyberpunk 2077

Cyberpunk 2077 news

Cyber Punk car





Less noise with more accurate results

A Closer Look at the Vehicles in Cyberpunk

By Keith Mitchell 3 days ago

The vehicles in Cyperpunk 2077 are not only varied, but there's a lot more in the game than we figured



(Image credit: CD Projekt Red)

We already knew that vehicles were included in Cyberpunk 2077, but we didn't know just how vast the selection would be. Sure, we imagined there would be a handful of vehicles, but our eyes were opened thanks to today's Cyberpunk 2077 Nightwire City Livestream, and we finally have a glimpse of just how many vehicles are going to be in the game.



Predict first, index later

Indexing only the **prospective user queries**



Indexing problem is turned into a **prediction problem**





Without massive data collection are we doomed to fail?



Data prediction

Use AI to predict the prospective user's query

Real User Queries	Generated User Queries		
game of thrones	game of thrones		
game of thrones cast	hbo game of thrones		
game of thrones season 8	game of thrones season 8		
game of thrones season 1	game of thrones tv		
game of thrones books	game of thrones cast		
game of thrones episodes	watch game of thrones		
watch game of thrones	game of thrones finale		
game of thrones season 5	game of thrones season 7		
game of thrones season 4	game of thrones john snow		
the mountain game of thrones	history behind game of thrones		



See all 304 videos »



Battle of the Bastards Jon and Sansa face Ramsav Bolton on the fields of Winterfell. Daenerys strikes back at her enemies. Theon and Yara arrive in

Tyrion advises Daenerys. Sansa forces Theon to tell her a secret. Cersei remains stubborn. Arya meets her first target. Jon and Tormund meet with the wildling elders.

Robb and Catelyn arrive at the Twins for the wedding. Jon is put to the test to see where his loyalties truly lie. Bran's group decides to split up. Daenerys plans an

A Salute to Women Directors



We take a look back to celebrate the pioneering women directors throughout cinematic history and their groundbreaking

Watch the video »

Related News





This solves the indexing problem

The next challenge is to maximize the recall and precision



It's all about the context

What is the **release date** of **Cyberpunk 2077** to **PS4** and **Xbox One**?

Google	What is the release date of Cyberpunk 2077 to PS4 and Xbox One? X Q Q All Videos Images Shopping News More Tools	۵ :::	С
	About 31,100,000 results (0.94 seconds) Cyberpunk 2077 / Initial release date December 10, 2020 After years of anticipation, CD Projekt released Cyberpunk 2077 for PlayStation 4, Stadia, Windows, and Xbox One on 10 December 2020, with PlayStation 5 and Xbox Series X/S versions planned to follow in 2021.	Cyberpunk 2077 < Video game 70% liked this video game Image: Coogle users Google users Image: Cyberpunk 2077 is an action role-playing video game developed and published by CD Projekt. The story takes place in Night City, an open world set in the Cyberpunk universe. Wikipedia	
	Cyberpunk 2077 - Wikipedia People also search for Assa The Ghost of Witcher of Ghost of The Dogs: Valhalla 3: Wild 2020 Yuld 2020 Y	Initial release date: September 17, 2020 Mode: Single-player video game Platforms: PlayStation 4, PlayStation 5, Xbox One, Google Stadia, Xbox Series X and Series S, Microsoft Windows Developers: CD Projekt, CD Projekt RED Designers: Mike Pondsmith, Konrad Tomaszkiewicz Composers: Marcin Przybyłowicz, Paul Leonard-	



Not an NLP problem

What is the **release date** of **Cyberpunk 2077** to **PS4** and **Xbox One**?

release date Cyberpunk 2077 PS4 Xbox One

boogle	release Xbox Cyberpunk date 2077 PS4 One	X 🎙 Q		🅸 🏼 🕻
	🔍 All 🖾 Images 🗉 News 🔗 Shopping 🕞 Videos ᠄ More	Tools		
	About 3,960,000 results (1.15 seconds)			
	Cyberpunk 2077 / Initial release date	10	Cyberpunk 2077	
	December 10, 2020	Contraction of the second	Video game	
			70% liked this video game Google users	
	People also search for		Cyberpunk 2077 is an action role-playing video game developed and published by CD Projekt. The story	
	Assassin's Creed Valhalla November 10		takes place in Night City, an open world set in the Cyberpunk universe. Wikipedia	
	2020		Initial release date: September 17, 2020	

What will you choose?

What is the **release date** of **Cyberpunk 2077** to **PS4** and **Xbox One**?

Cyberpunk 2077

PS4 console





Contextual Graph



Taxonomy vs. Contextuality



Contextual graph





Contextual graph





Graph statistics





Data collection

What to extract

How to connect



Scalable architecture







Brain inspired **auto associative networks** and **statistical pattern recognition**



Applications to Search



Possible query forms

gta	Q Sentire Web	
ALL IMAGES NEWS		
110,769 Results	Contextual Graph (i)	
Grand Theft Auto: The Official Site http://www.rockstargames.com/grandtheftauto/ Grand Theft Auto V. Out now for PlayStation4, Xbox One, PlayStation3, Xbox 360, and PC. A bold new direction in open-world freedom, storutelling, mission-based gamenlay, and opling multiplayer. Grand The	grand theft auto	70.02%
	grand theft auto v	4.28%
GTA - WIKIPEdia https://en.wikipedia.org/wiki/GTA GTA may refer to:.	gta san andreas	3.76%
Crand Thaft Auto V kapan Microsoft Store nl NI	gta v	3.61%
https://www.microsoft.com/nl-nl/p/grand-theft-auto-v/bpj686w6s0nh? When a young street hustler, a retired bank robber and a terrifying psychopath find themselves entangle some of the most frightening and deranged elements of the criminal underworld, the U.S. government a	ed with and	2 53%
Grand Theft Auto V on Reddit		
https://www.reddit.com/t/grand_theft_auto_v/ Find the best posts and communities about Grand Theft Auto V on Reddit. https://github.com/LukeRoss00/gta5-real-mod/releases Please see the READMEfor important instruction	ions X Greate	er Toronto Area
Grand Theft Auto - VICE	• •	
https://www.vice.com/en/topic/grand-theft-auto		

Everything with the topic 'Grand Theft Auto' on VICE. Rockstar Games and Take-Two Interactive obtained a

Possible query forms

gone wind	Q Entire Web \checkmark	
ALL IMAGES NEWS		
163,342 Results	Contextual Graph (i)	
Gone with the Wind - Wikipedia	gone with the wind	100%
Gone with the Wind may refer to:.		
Gone with the Wind - NYT Watching		
https://www.nytimes.com/watching/titles/gone-with-th	ne-wind	
null. null.		
Gone With The Wind GIFs - Find & Share on	GIPHY	
null. null. Gone With The Wind GIFs - Find & Share on https://giphy.com/search/gone-with-the-wind/	GIPHY	
null. null. Gone With The Wind GIFs - Find & Share on https://giphy.com/search/gone-with-the-wind/ Find GIFs with the latest and newest hashtags! Search, discover and GIFs.	GIPHY d share your favorite Gone With The Wind	
null. null. Gone With The Wind GIFs - Find & Share on https://giphy.com/search/gone-with-the-wind/ Find GIFs with the latest and newest hashtags! Search, discover and GIFs. Gone With The Wind The Marquee Blog - CN	GIPHY d share your favorite Gone With The Wind IN.com Blogs	
null. null. Gone With The Wind GIFs - Find & Share on https://giphy.com/search/gone-with-the-wind/ Find GIFs with the latest and newest hashtags! Search, discover and GIFs. Gone With The Wind The Marquee Blog - CN http://marquee.blogs.cnn.com/category/movies/gone-	GIPHY d share your favorite Gone With The Wind IN.com Blogs -with-the-wind/	
null. null. Gone With The Wind GIFs - Find & Share on https://giphy.com/search/gone-with-the-wind/ Find GIFs with the latest and newest hashtags! Search, discover and GIFs. Gone With The Wind The Marquee Blog - CN http://marquee.blogs.cnn.com/category/movies/gone- The CNN.con Marquee Blog: The spot for breaking entertainment no the inside story on whats happening in Hollywood. The CNN.con Margue	GIPHY d share your favorite Gone With The Wind IN.com Blogs -with-the-wind/ ews, recaps of your favorite TV shows and arquee Blog: The spot for breaking	
null. null. Gone With The Wind GIFs - Find & Share on the https://giphy.com/search/gone-with-the-wind/ Find GIFs with the latest and newest hashtags! Search, discover and GIFs. Gone With The Wind The Marquee Blog - CN http://marquee.blogs.cnn.com/category/movies/gone- The CNN.con Marquee Blog: The spot for breaking entertainment not the inside story on whats happening in Hollywood. The CNN.con Marguee Blog - CNN.con Marguee Blog - CNN.con Marguee Blog: The spot for breaking entertainment not the inside story on whats happening in Hollywood. The CNN.con Marguee Blog - CNN.con Marguee Blog - CNN.con Marguee Blog: The spot for breaking entertainment not be inside story on what happening in Hollywood. The CNN.con Marguee Blog - CNN.con Marguee Blog - CNN.con Marguee Blog: The Spot for breaking entertainment not be inside story on what happening in Hollywood. The CNN.con Marguee Blog - CNN.con Marguee Blog - CNN.con Marguee Blog - CNN.con Marguee Blog: The Spot for breaking entertainment not be inside story on what happening in Hollywood. The CNN.con Marguee Blog - CNN.con Marguee Blog - CNN.con Marguee Blog - CNN.con Marguee Blog: The Spot for breaking entertainment not be inside story on what happening in Hollywood. The CNN.con Marguee Blog - CNN.con Margueee Blog - CNN.con Margueeee Blog - CNN.con Margueeee Blog - CNN.con Margueeee Blog - CNN.con Margueeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeee	GIPHY d share your favorite Gone With The Wind IN.com Blogs -with-the-wind/ ews, recaps of your favorite TV shows and arquee Blog: The spot for breaking	
null. null. Gone With The Wind GIFs - Find & Share on the https://giphy.com/search/gone-with-the-wind/ Find GIFs with the latest and newest hashtags! Search, discover and GIFs. Gone With The Wind The Marquee Blog - CN http://marquee.blogs.cnn.com/category/movies/gone- The CNN.con Marquee Blog: The spot for breaking entertainment not the inside story on whats happening in Hollywood. The CNN.con Mar Gone With The Wind : NPR https://www.npr.org/tags/135987243/gone-with-the-w	GIPHY d share your favorite Gone With The Wind IN.com Blogs -with-the-wind/ ews, recaps of your favorite TV shows and arquee Blog: The spot for breaking	

Query augmentation

Google	dawn ps4	IJ	dawn ps4 Q	💿 Er	tire Web	\sim
	Q All 🖬 Images Videos ⊘ Shopping 🗉 News : More		ALL IMAGES NEWS			
	About 47,700,000 results (0.72 seconds)		312,534 Results Contextual Graph (i)			
	Sony Until Dawn Rush Of Blood V (PS4) from Sony at best Buy Sony Until Dawn Rush Of Blood V (PS4) from Sony online at best deals only enjoy √no cost EMI & √free shipping.		Buy Now Horizon Zero Dawn PS4 Games https://www.playstation.com/en-gb/games/horizon-zero Horizon Zero Dawn: Complete Edition Horizon Zero Dawn: Complet ray Disc The Frozen Wilds expansion The Art of Horizon Zero Dawr until dawn ps4	22.44%	-	
	https://www.smythstoys.com > > PlayStation 4 Games ▼ Horizon Zero Dawn Complete Edition PS4 - Smyths Toys Buy Horizon Zero Dawn Complete Edition PS4 Online at Smyths Toys Ireland Or (Local Smyths Toys! ✔ Click & Collect Available ✔ FREE ★★★★★ Rating: 5 - 22 reviews · €19.99 · In stock		Until Dawn - Wikipedia https://en.wikipedia.org/wiki/Until_Dawn Until Dawn is a horror adventure video game developed by Supermassive Games and published by Sony Computer Entertainment for PlayStation 4. It was originally scheduled to be released for PlayStation 3 and	12.66%	— playstation 4	new dawn
	https://store.ubi.com → → Eranchises A-P → Far Cry ▼ Buy Far Cry New Dawn Standard Edition for PS4, Xbox C Buy Far Cry New Dawn Standard Edition for PS4, Xbox One and PC at the Officia and dive into a transformed, vibrant, post-apocalyptic Hope		Far Cry New Dawn - IGN http://ign.com/far-cry-new-dawn The Frozen Wilds is a great 15-hour experience that reminded me just how wonderful this world is. The Frozen Wilds is a great 15-hour experience that reminded me just how wonderful this world is.		until down po	1 2000
	https://www.ign.com > articles > play-at-home-playstati • Play at Home: PlayStation Offering 10 More Games, Inclu 17 Mar 2021 — Sony is offering Horizon: Zero Dawn and nine other games for free as part of its Play at Home initiative.		Horizon Zero Dawn - GameSpot http://gamespot.com/horizon-zero-dawn/ Horizon Zero Dawn is an action role playing game exclusively for the PlayStation 4. As Horizon Zero Dawns main protagonist Aloy, a skilled hunter, explore a vibrant and lush world inhabited by mysterious. PC			
	https://www.takealot.com>Gaming>Games ▼ Until Dawn (PS4) Buy Online in South Africa takealot.cc		Horizon: Zero Dawn Archives DualShockers https://www.dualshockers.com/tag/horizon-zero-dawn/			

Contextual Graph It's all about the context

