Under the Hood: Building posts search

October 24, 2013 at 10:58am

Last week we added the ability to search posts using Graph Search, a feature that has been two years in the making. With one billion new posts added every day, the posts index contains more than one trillion total posts, comprising hundreds of terabytes of data. Indexing these posts and building a system to return real-time results has been a significant engineering challenge—and this is just beginning.

Collecting the Data

One of the biggest challenges we faced in building this product was that Facebook's underlying data schema reflects the needs of a rapidly iterated web service. New features often demand changes to data schemas, and our culture aims to make these changes easy for engineers. However, these variations makes it difficult to sort posts by time, location, and tags as wall posts, photos, and check-ins all store this information differently. We currently have 70 different kinds of data we sort and index on, many of them specific to certain types of posts. In addition, the data is contained in a production MySQL database. This means that harvesting this data puts a significant load on databases that are simultaneously serving production traffic, and the process must be carefully monitored.

Building the Index

We store our harvested data in an HBase cluster, from which we execute Hadoop map-reduce jobs to build our index in a highly parallel process. This index-building process converts the raw data into a search index that works with Unicorn, our search infrastructure. We separate the data into two parts - the document data and the inverted index. The document data for each post contains information that will later be used for ranking results. The inverted index contains what is traditionally considered to be a search index, and building the inverted index requires going through each post and determining which hypothetical search filters match.

Updating the Index

To update the index, we subscribe to changes in the MySQL databases using a technology called Wormhole. Whenever a new post is created, an existing post is modified, a post is deleted, or some connection to a post is edited, we schedule an update on the corresponding post. To avoid code duplication, we run these updates through the same logic mentioned in the "Collecting the Data" section above. However, there is one major distinction: when we originally collect the data, we intentionally bypass our caches, as we want to avoid sending them requests for old data that is likely missing from the cache. When processing an update to the index we hit those caches, as we expect the data to have been recently accessed and be present in the cache.

Serving the Index

The posts index is much larger than other search indexes that Facebook maintains. Before we started working on the ability to search through posts, all Facebook search indexes were served entirely from RAM. This was ideal for quick lookup and was a tenable setup given reasonably small search indexes. However, storing more than 700 terabytes in RAM imposes a large amount of overhead, as it involves maintaining an index that is spread across many racks of machines. The performance cost of having these machines coordinate with each other drove the Unicorn team to look into new solutions for serving the posts index. The solution we decided on involves storing the majority of the index on solid-state flash memory. We managed to preserve performance by carefully separating out the most frequently accessed data structures and placing those in RAM.

Ranking Results

With a trillion posts in the index, most queries return many more results than anyone could ever read. This leads us to the results ranking step. To surface content that is valuable and relevant to the user, we use two primary techniques: query rewriting and dynamic result scoring. Query rewriting happens before the execution of the query, and involves tacking on

optional clauses to search queries that bias the posts we retrieve towards results that we think will be more valuable to the user. Result scoring involves sorting and selecting documents based on a number of ranking "features," each of which is based on the information available in the document data. In total, we currently calculate well over a hundred distinct ranking features that are combined with a ranking model to find the best results. We will continue to work on refining these models as we roll out to more users and listen to feedback.

Project History
Like many other products at Facebook, the ability to search over posts was originally conceived as a Hackathon project. My second day as a Facebook intern coincided with a company-wide Hackathon, and I spent the night aiming to implement a way for my friends and me to find old posts we had written. I quickly discovered that the project was much more challenging than I had first anticipated. However, the engineering culture at Facebook meant that I was supported and encouraged to continue working on it, despite the scope of the project. The majority of the work--infrastructure, ranking, and product--has been accomplished in the past year by a few dozen engineers on the Graph Search team. We are very excited to share the results of our hard work with people using Facebook, and look forward to seeing how people use the new feature and improving it based on their feedback. We hope that being able to search for posts will enable a richer Facebook experience for everyone.

Ashoat Tevosyan is an engineer on the search quality and ranking team.

Like · Comment · Share

Seung−Hwan Lim, Crear web de cupones, Sufian Dar and 394 others like this.   Most Relevant

100 shares

André Lima A big challenge sliced in five smaller challenges. Even though it took two years. Great history. Congrats to have accomplished and to share the experience.
3 · October 24, 2013 at 1:24pm

Raghav Channarayappa Try giving a search for posts on timeline, pages, on groups first..
3 · October 24, 2013 at 11:30am

3 Replies

Andrew A Blair Great work!
2 · October 24, 2013 at 11:13am

Eric Man Slick.
2 · October 24, 2013 at 11:02am

Karthik Krishnan Excited! Waiting to have this update. Great work from the team, kudos! Any timeline on when it would be available to all?
Here are some of my requests based on playing with photo search so far:... See More
1 · October 25, 2013 at 9:25pm

Mike Nguyen Why do you guys never reply or answer my complaints especially when I spent money for marketing feeds I feel like you guys do not respect the users.
October 24, 2013 at 3:01pm

Jacob Andresen Sounds like a fun challenge! Do you have live updates for the posts index? It could be fun to create an app that aggregates live post subjects for communities and groups.
1 · October 24, 2013 at 2:17pm

1 Reply

Shalif Mayans ok
1 · October 24, 2013 at 11:06am

Susan Ditler−Brown Help! Is anybody else getting really old posts on their feed and almost no recent posts? FB messing with us again? When I say old, I mean like a week or 2 old.. If I force refresh it shows lots of new post numbers but not the posts. There are a couple ... See More
2 · October 24, 2013 at 2:52pm

Raghav Channarayappa There is even more amount of data Google indexes on a daily basis.. What do u have to say for that?
1 · October 24, 2013 at 11:28am

Margie Hearron It's 2014 and we still don't have POST SEARCH. We can only hashtag search on desktop FB and the mobile web version of FB in a browser. Why the delay on post search, Facebook Android Beta Facebook for iPhone???
July 28, 2014 at 2:39pm

Kevin Von Lethe Any news ??
May 21, 2014 at 3:27pm

John Peter Thompson Garcés How long does it take to index all of the posts on Facebook?
May 15, 2014 at 5:57pm

Dan Kegel I still don't have this feature. No way to search for posts. It's really frustrating to see lots of articles saying "Facebook improves search!" when they seem to have actually made it worse for me.
April 13, 2014 at 7:46pm

Dan Kegel Has the rollout been halted? I don't seem to have this feature, even though I have the rest of graph search. Searching for posts about "shoes" yields the error "This search isn't available to you yet". I've run into several other people who have the same problem.
February 16, 2014 at 7:37pm · Edited

임남규 I've ever not searched a certain post while I'm using my facebook account. Is this post now meaning that my account have configured a hidden query comprised of my facebook activity to deliver to inverted index ??
November 4, 2013 at 9:50pm

Khulekani Mjey Khumalo y u guys dont giv us job
November 3, 2013 at 8:29am

Becky Wludyka How do you delete your facebook page
October 28, 2013 at 6:47am

Jason Kim Yo, Ashoat! Creeping on people become more efficient!   Thanks bro. In Alexis we trust
October 25, 2013 at 6:43pm

Kesavsundar Gopalakrishnan Awesome !
October 25, 2013 at 9:11am

Steve Bradley You used to be able to do that by default. I could enter a word like "halloween" and the results would show where it appeared on pages as well as people's posts. I think the option was "view posts by others".
October 24, 2013 at 10:18pm

Sandeep Mangal Hats-off to u guys..did a grt job..but u gus uses NOSQL nd here u implementd on MYSQL??
October 24, 2013 at 11:12am

David Jenkins I have good friends Facebook.Com.
1 · October 24, 2013 at 11:05am

Mohamed A. Baset Waiting
January 23, 2014 at 6:24pm

Aman Singh like this cool new page about technology
https://www.facebook.com/pages/Computers-Bytes-n-Fun/283073718500520?hc_location=stream
October 28, 2013 at 9:37pm

Raghav Channarayappa On facebook I mean not on Google search, Tom Jackson
October 25, 2013 at 10:03am · Edited

Rohit Pathak I had the idea that facebook used NoSql
October 24, 2013 at 7:18pm

Sonal Barabde If u r a mechanical engineer then must like this page try it friendz Knowledge Is Not Only For The Toppers"
October 24, 2013 at 11:00pm

Kaiypov Emil ANTI-MILITARY "SPAM" Dear Facebook Engineering Team
https://www.facebook.com/Engineering! Please support Global truce - Pause all wars anti-war initiative! Best wishes from Bishkek, Kyrgyzstan!!! Launch of the First Anti-military satellite. ...
See More

Facebook Engineering
Computers/Technology

1,282,089 Likes
3,373 talking about this

January 8 at 1:48pm

Belajar Bahasa Inggris   Kucing Lucu / Funny Cats
http://a1-weblog-9z.blogspot.com/.../Unit-Link-Terbaik-Di...
1 · March 16, 2014 at 4:53am

UNETE Porfavor A Hola Amigos Esta Es Una Nueva Pagina Se Buscan Administradores Solo Unete Y Lo Seras
https://www.facebook.com/.../Puta-Madre.../1400519360183381
October 26, 2013 at 4:23pm

Sign Up   Log In   Messenger   Mobile   Find Friends   Badges   People   Pages   Places   Games
Locations   About   Create Ad   Create Page   Developers   Careers   Privacy   Cookies   Terms   Help

Facebook © 2015

https://www.facebook.com/note.php?note_id=10151755593228920   3/4

English (US)