

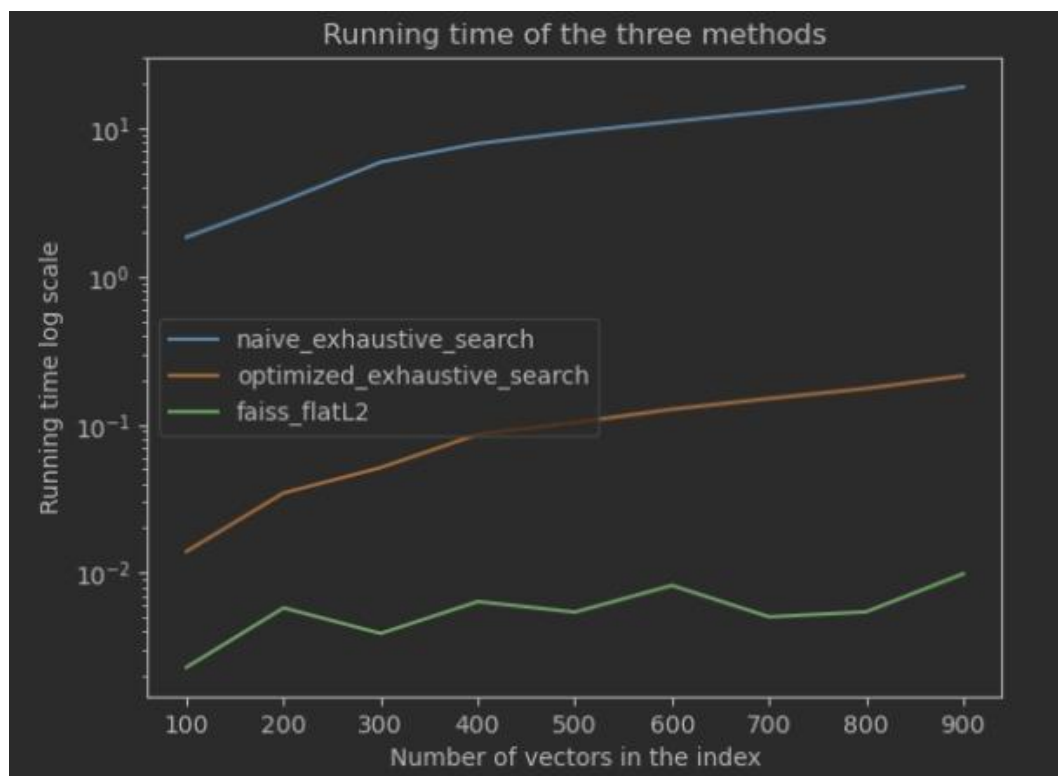
HW1-Vector Search

Submitted by: Roy Azuelos 205736879, Emily Ramim 209606144 Yonatan Levin
213125164

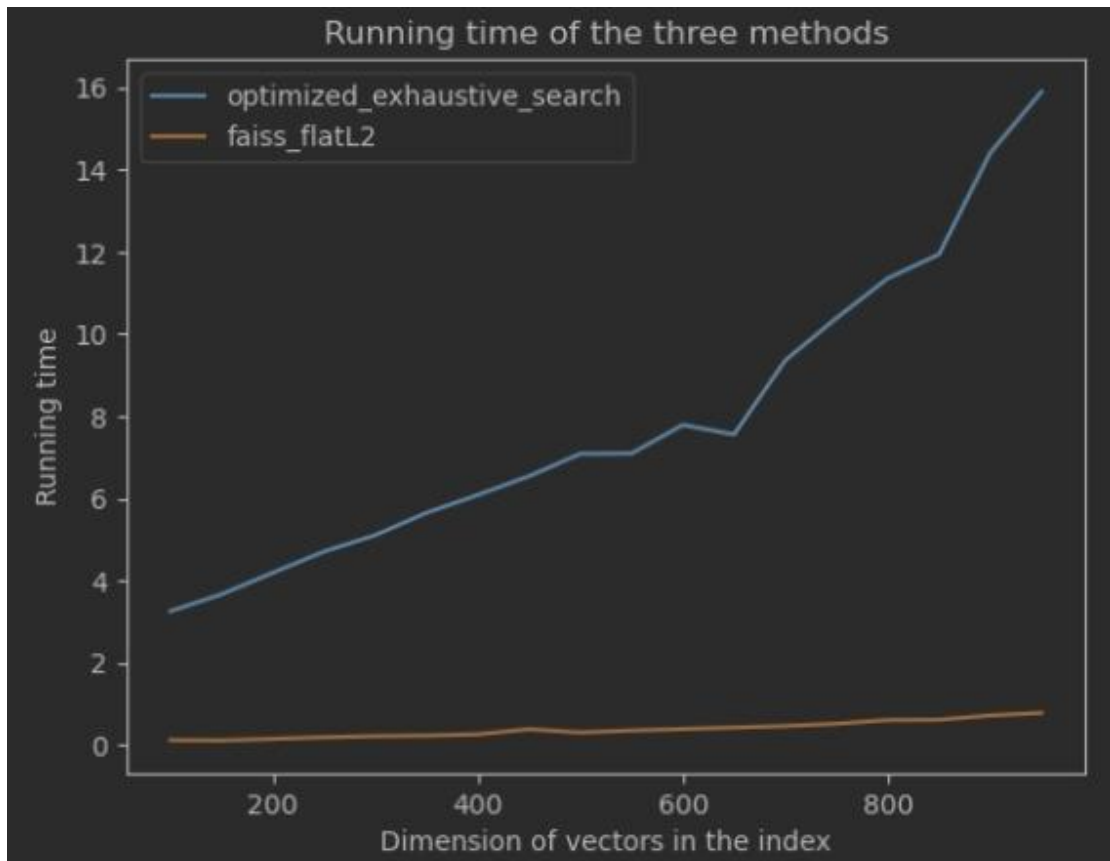
[Github repository](#)

Part 1-Faiss

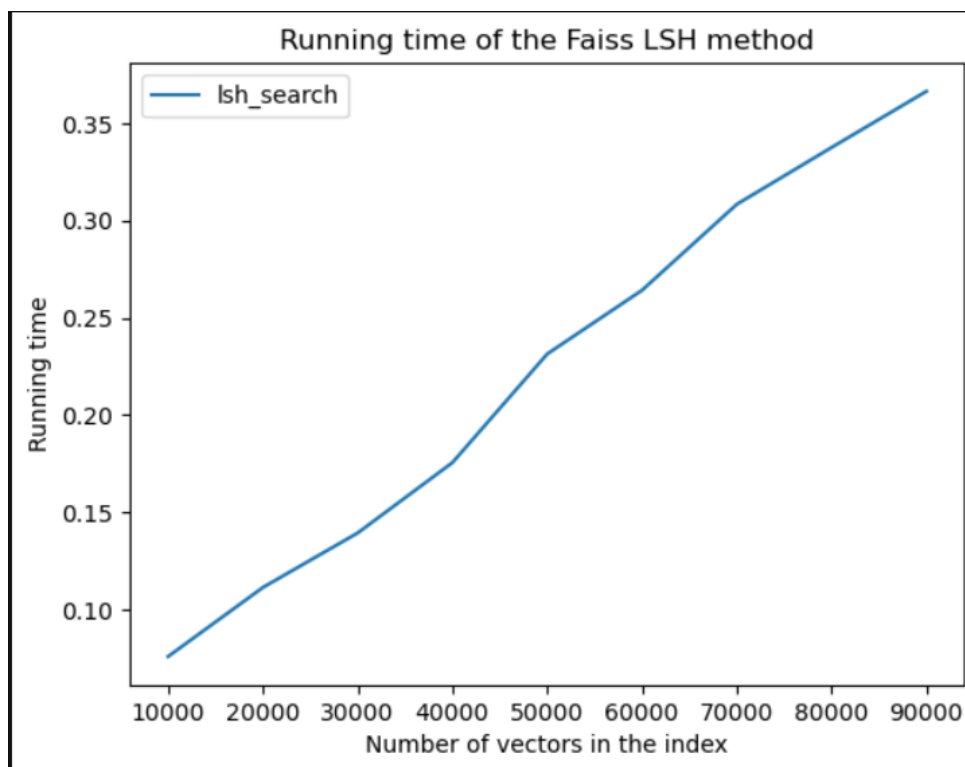
displaying the running time of the three methods as a function of the
number of vectors in the index:



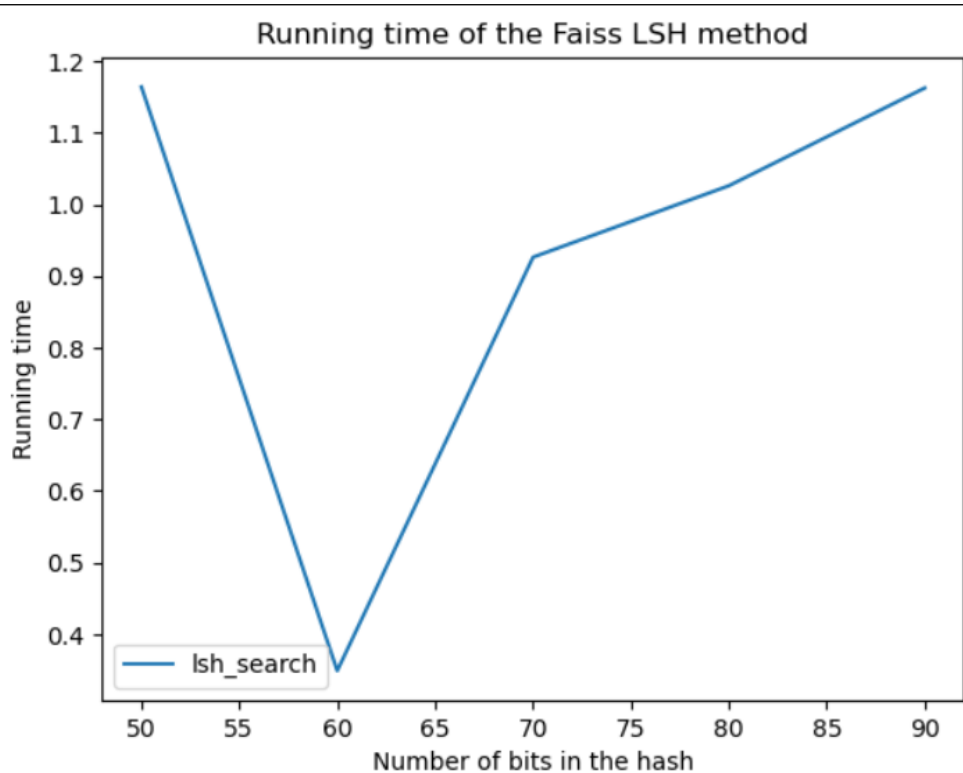
displaying the running time of the three methods as a function of the
dimensionality of the vectors:



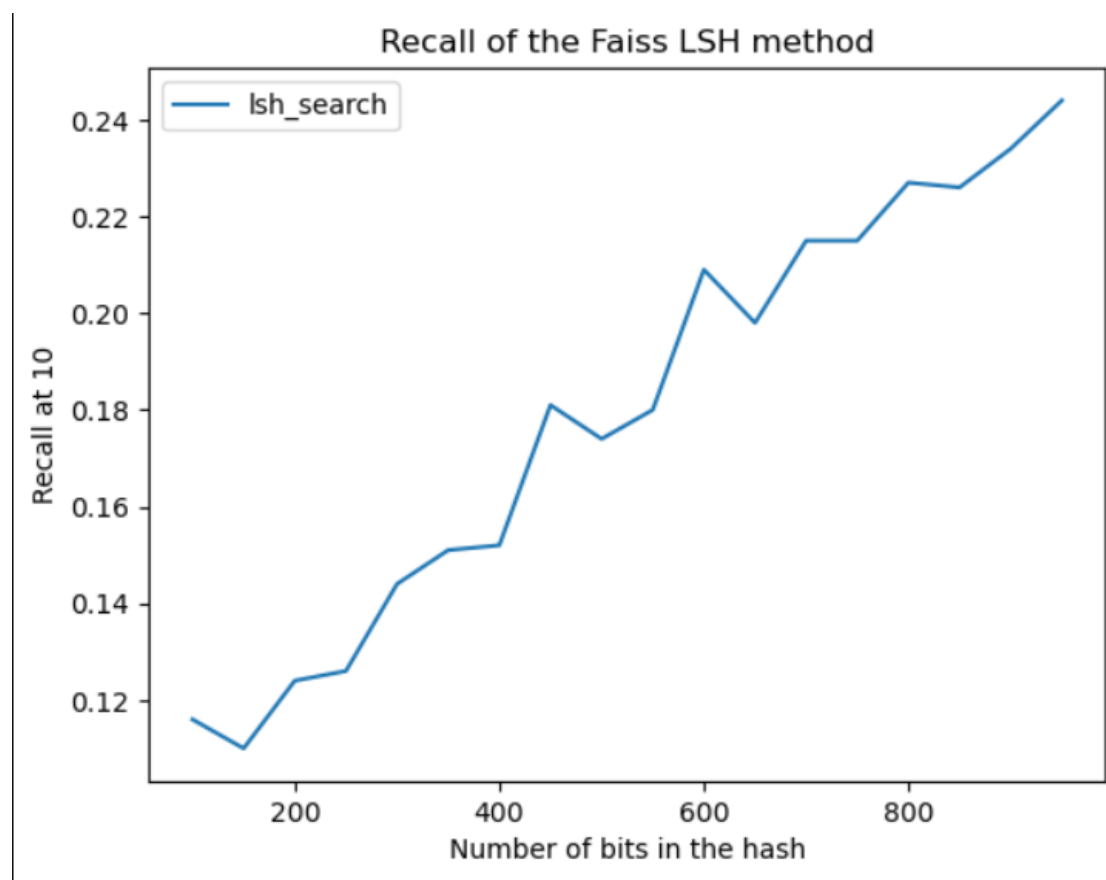
A plot displaying the running time of Faiss LSH as a function of the number of vectors in the index:



.A plot displaying the running time of Faiss LSH as a function of nbits



A plot displaying the recall@k of Faiss LSH as a function of nbits:



Part 3 -Pinecone VectorDB and RAG:

- The dataset that we have decided to use is '[wikitext-2-raw-v1](#)'. We chose this dataset due to the large size of it (around 44k items), most items are relatively short (less than 400 words), the dataset is static, and it contains data about various topics. It also contains data about very specific domains, many of which aren't common on the web. Because of that, many of the LLMs available don't contain information about those domains, and they are prone to hallucinations.
- For the query: ""When did Raphael Tuck buy four of Barker's 'little drawings'?", we got those results:

```
Original response:
Raphael Tuck bought four of Barker's "little drawings" in 1866.
-----
Augmented response:
In 1911, Raphael Tuck & Sons bought four of Barker's "little drawings."
-----
Source knowledge:
In 1911, Raphael Tuck & Sons bought four of Barker's "little drawings" for half a sovereign, and published them as postcards. In October 1911, she won second prize in the Croydon Art Society's poster competition, and shortly afterward was elected the youngest member of the Society. The art critic for the Croydon Advertiser remarked, "Her drawings show a remarkable freedom of spirit. She has distinct promise."

Following her father's death in June 1912, the seventeen-year-old Barker submitted art and poetry to My Magazine, Child's Own, Leading Strings, and Raphael Tuck annuals in an effort to support both her mother and sister. Her sister Dorothy taught kindergarten in two private schools before opening a kindergarten at home. She brought in some money for the family's support while supervising the household.

Barker was equally proficient in watercolour, pen and ink, oils, and pastels. Kate Greenaway and the Pre-Raphaelites were the principal influences on her work. She claimed to paint instinctively and rejected artistic theories. Barker died in 1973. Though she published Flower Fairy books with spring, summer, and autumn themes, it wasn't until 1985 that a winter collection was assembled from her remaining work and published posthumously.
```

As we can see, while the original response gave us false information (according to the source knowledge), the augmented response was more akin to the truth.

For the query "Who sang the opening theme in Valkyria Chronicles III?" we got this responses:

```
Original response:
*Valkyria Chronicles III* is an amazing game with a beautiful opening theme! The song is called "Blue Star" and it is performed by Japanese singer and voice actress, Kanako Kotera.
-----
Augmented response:
May'n
-----
Source knowledge:
The game began development in 2010, carrying over a large portion of the work done on Valkyria Chronicles II. While it retained the standard features of the series, it also underwent multiple adjustments, such as making the game more forgiving for series newcomers. Character designer Raita Honjou and composer Hitoshi Sakimoto both returned from previous entries, along with Valkyria Chronicles II director Takeshi Ozawa. A large team of writers handled the script. The game's opening theme was sung by May'n.

= Valkyria Chronicles III =

The music was composed by Hitoshi Sakimoto, who had also worked on the previous Valkyria Chronicles games. When he originally heard about the project, he thought it would be a light tone similar to other Valkyria Chronicles games, but found the themes much darker than expected. An early theme he designed around his original vision of the project was rejected. He redid the main theme about seven times through the music production due to this need to reassess the game. The main theme was initially recorded using orchestra, then Sakimoto removed elements such as the guitar and bass, then adjusted the theme using a synthesizer before redoing segments such as the guitar piece on their own before incorporating them into the theme. The rejected main theme was used as a hopeful tune that played during the game's ending. The battle themes were designed around the concept of a "modern battle" divorced from a fantasy scenario by using modern musical instruments, constructed to create a sense of atonality. While Sakimoto was most used to working with synthesized music, he felt that he needed to incorporate live instruments such as orchestra and guitar. The guitar was played by Mitsuhiro Ohta, who also arranged several of the later tracks. The game's opening song, "If You Wish for..." (もしも君が願うのなら, Moshimo Kimi ga Negauno Nara), was sung by Japanese singer May'n. Its theme was the reason soldiers fought, in particular their wish to protect what was precious to them rather than a sense of responsibility or duty. Its lyrics were written by Seiko Fujibayashi, who had worked on May'n on previous singles.
```

As we can see from this example, the original response from the model gave us incorrect answers, while also giving us unnecessary info about it's "opinions" about the song. After using the RAG pipeline, and after we made chunking to the dataset, the new answer was correct and exactly to the point of the query.

For our final example, we asked "When did Columbus have the best chance of receiving the first overall pick?". We got those answers:

```
Original response:
2015
-----
Augmented response:
Columbus had the best chance of receiving the first overall pick in the 2012 NHL Entry Draft lottery.
-----
Source knowledge:
Finishing with the worst record in the NHL , Columbus had the best chance of receiving the first overall pick in t
he 2012 draft . With the NHL 's weighted draft lottery the Blue Jackets had a 48 @-@ 2 % chance of drafting first o
verall . However , the lottery was won by the Edmonton Oilers , who proceeded to leapfrog Columbus and secure the n
umber one draft pick for a third consecutive year . It was the fifth time that the Blue Jackets were dropped one dr
aft position in the franchise 's 12 lottery participations .

Two weeks prior to the NHL trade deadline , Columbus announced that unlike earlier in the season , they would list
en to trade proposals involving Rick Nash , though they were not actively shopping him . Howson stated that the tea
m was open to all options for improving the team , including trading Nash . Speculation was that in return for Nash
the Blue Jackets would ask for a " combination of young , proven players , high @-@ end prospects and draft picks .
" Leading up to the trade deadline , the Blue Jackets dealt Antoine Vermette to the Phoenix Coyotes for two draft p
icks and goaltender Curtis McElhinney . Despite being injured at the time , the acquisition of McElhinney was belie
ved to give Columbus the flexibility to trade Curtis Sanford . The following day , on February 23 , Columbus traded
Jeff Carter to the Kings . In the deal , Columbus acquired defenseman Jack Johnson and a first @-@ round draft pick
; the team was given the choice of taking the pick in either 2012 or 2013 . At the deadline , Columbus was unable t
o come to terms on a deal involving Nash , but they did make one more move ; they sent center Samuel Pahlsson to th
e Vancouver Canucks in exchange for two fourth @-@ round draft picks and minor league defenseman Taylor Ellington .
Following the trade deadline , Howson announced that the team had attempted to trade Nash at the player 's request
. Nash stated that he had requested the trade after being informed that the franchise was going into another rebuil
ding phase . He further noted that he felt that he " could be a huge part of that towards bringing assets in , " an
d in his view " it was the best thing for the team , the organization , and personally for [ his ] career . " After
the personnel changes , the Blue Jackets closed out the month with a three @-@ game losing streak .
```

Similarly to the first 2 examples, the original response gave us completely incorrect answers. However, here the original model gave us a shorter and to the point answer, despite the fact that it was wrong.

- In our code, we planned the retrieval system so that given the query, the system will select the top 3 results from the knowledge base. We used the 'all-MiniLM-L6-v2' embedding model , with token chunking(for each row in the dataset) in order to encode the dataset. We also created a Pinecone vector database using the cosine metric system. We then updated the Pinecone index by training the model with the knowledge base.
- As we can see from the results, while the RAG pipeline did not managed to completely get rid of the hallucinations in some of the queries, there was some improvement in the answers that were given, and if the model didn't knew the answer, it simply

said so rather than fabricated an answer as it did in the original QA model. Another insight is that the RAG pipeline model gave us short and to-the-point answers to our queries, rather than give us long and excruciating to read answer.