

Davide Lagano
Georgios Patrikis
Björkeröth, Didrik
Eliasson, Robin

Assignment Group 03

Assignment 1: Quality of Retrieval by Major Search Engines

We choose to investigate the following search engines:

- Google
- Bing
- DuckDuckGo

Information needs

Information need 1: “Any news about the war in Syria?”

Query 1: “News war+Syria”

The research considered 37 links as relevant. The list of links can be found at the appendix.

Information need 2: “You are burned out and you are looking for psychological advice for highly stressed people.”

Query 2: “psychological advice+burned out+relieve stress”

The research considered 26 links as relevant. The list of links can be found at the appendix.

Information need 3: “You want to find out how hashing and hash tables work.”

Query 3: “How hashing+hashtables work”

The research considered 45 links as relevant. The list of links can be found at the appendix.

Information need 4: “You are looking for information on the history of La Liga.”

Query 4: history of+la liga

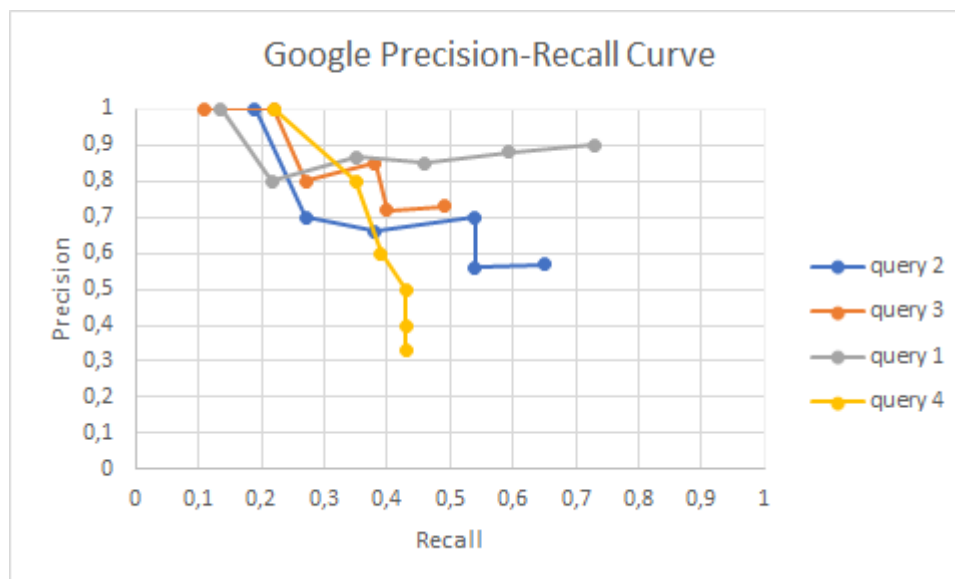
The research considered 23 links as relevant. The list of links can be found at the appendix.

For each query in every search engine, we consider relevant the firsts 30 results. For this reason, we go through 90 websites and we choose the relevant results.

Precision-Recall

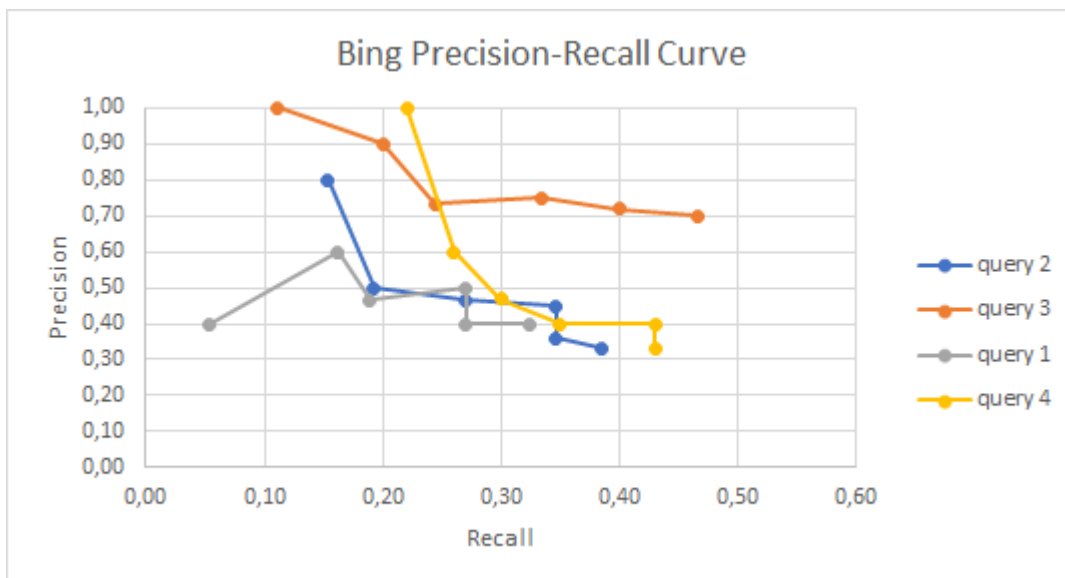
Search engine 1:

	Google							
	Query 1		Query 2		Query 3		Query 4	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
Top 5	5/5 = 1	5/37 = 0.14	5/5 = 1	5/26 = 0,19	5/5=1	5/45 = 0.11	1	0.22
Top 10	8/10 = 0.8	8/37 = 0.22	7/10 = 0.7	7/26 = 0.27	10/10 = 1	10/45 = 0.22	0.8	0.35
Top 15	13/15 = 0.87	13/37 = 0.35	10/15 = 0.66	10/26 = 0.38	12/15 = 0.8	12/45 = 0.27	0.6	0.39
Top 20	18/20 = 0.9	18/37 = 0.49	14/20 = 0.7	14/26 = 0.54	17/20 = 0.85	17/45 = 0.38	0.5	0.43
Top 25	23/25 = 0.92	23/37 = 0.62	14/25 = 0.56	14/26 = 0.54	18/25 = 0.72	18/45 = 0.4	0.4	0.43
Top 30	27/30 = 0.9	27/37 = 0.73	17/30 = 0.57	17/26 = 0.65	22/30 = 0.73	22/45 = 0.49	0.33	0.43



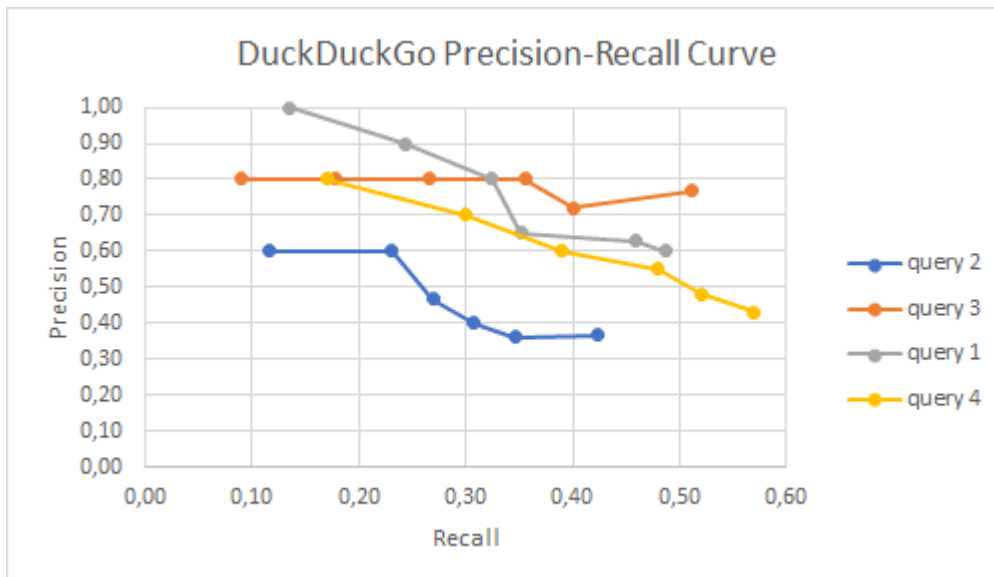
Search engine 2:

	Bing							
	Query 1		Query 2		Query 3		Query 4	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
Top 5	$2/5 = 0.4$	$2/37 = 0.05$	$4/5 = 0.8$	$4/26 = 0.15$	$5/5 = 1$	$5/45 = 0.11$	1	0.22
Top 10	$6/10 = 0.6$	$6/37 = 0.16$	$5/10 = 0.5$	$5/26 = 0.19$	$9/10 = 0.9$	$9/45 = 0.2$	0.6	0.26
Top 15	$7/15 = 0.47$	$7/37 = 0.19$	$7/15 = 0.47$	$7/26 = 0.27$	$11/15 = 0.73$	$11/45 = 0.24$	0.47	0.3
Top 20	$10/20 = 0.5$	$10/37 = 0.27$	$9/20 = 0.45$	$9/26 = 0.35$	$15/20 = 0.75$	$15/45 = 0.33$	0.4	0.35
Top 25	$10/25 = 0.4$	$10/37 = 0.27$	$9/25 = 0.36$	$9/26 = 0.35$	$18/25 = 0.72$	$18/45 = 0.4$	0.4	0.43
Top 30	$12/30 = 0.4$	$12/37 = 0.32$	$10/30 = 0.33$	$10/26 = 0.38$	$21/30 = 0.7$	$21/45 = 0.47$	0.33	0.43



Search engine 3:

	DuckDuckGo							
	Query 1		Query 2		Query 3		Query 4	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
Top 5	5/5 = 1	5/37 = 0.14	3/5 = 0.6	3/26 = 0.11	4/5 = 0.8	4/45 = 0.09	0.8	0.17
Top 10	9/10 = 0.9	9/37 = 0.24	6/10 = 0.6	6/26 = 0.23	8/10 = 0.8	8/45 = 0.18	0.7	0.3
Top 15	12/15 = 0.8	12/37 = 0.32	7/15 = 0.47	7/26 = 0.27	12/15 = 0.8	12/45 = 0.27	0.6	0.39
Top 20	13/20 = 0.65	13/37 = 0.35	8/20 = 0.4	8/26 = 0.31	16/20 = 0.8	16/45 = 0.36	0.55	0.48
Top 25	17/25 = 0.68	17/37 = 0.46	9/25 = 0.36	9/26 = 0.35	18/25 = 0.72	18/45 = 0.4	0.48	0.52
Top 30	18/30 = 0.6	18/37 = 0.49	11/30 = 0.37	11/26 = 0.42	23/30 = 0.77	23/45 = 0.51	0.43	0.57



Precision and recall formula

The following formula was used to calculate precision and recall:

Precision = $TP / TP + FP$ (TP = The correct documents retrieved, FP = The wrong documents retrieved).

Recall = $TP / TP + FN$ (FN = Documents missed by one search engine, but collected by at least another one).

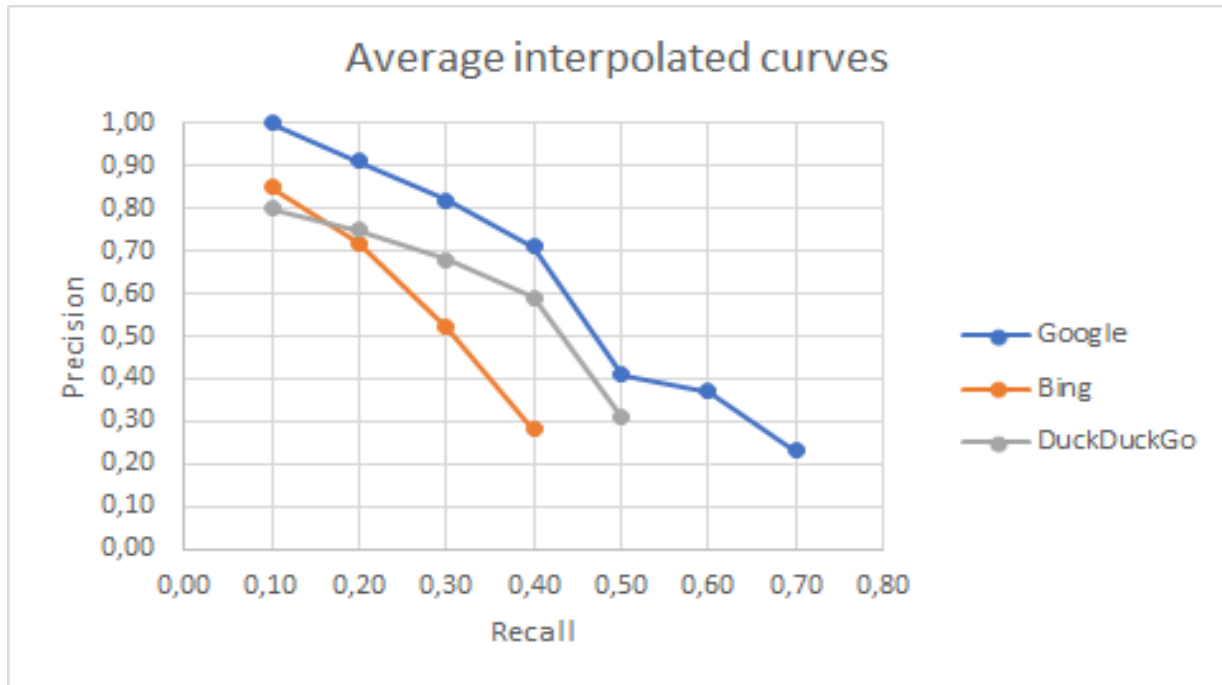
Interpolated precision

To calculate the interpolated precision, we used the following standard recall values: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 and 1. The greatest noted precision value to the right of this standard recall is what decides what the interpolated precision is. For example if the “line” in the graph is ascending and never dropping, then the interpolated precision value will be the noted precision furthest to the right. Also of note is that if the recall value hit dead on for the noted precision, and there were no greater precision value to the right, then this value becomes the interpolated value.

Google interpolated precision					
Standard recall	Query 1	Query 2	Query 3	Query 4	Average
0.1	1	1	1	1	1
0.2	0.92	0.7	1	1	0.91
0.3	0.92	0.7	0.85	0.8	0.82
0.4	0.92	0.7	0.73	0.5	0.71
0.5	0.92	0.7	0	0	0.41
0.6	0.92	0.57	0	0	0.37
0.7	0.9	0	0	0	0.23
0.8	-	-	-	-	-
0.9	-	-	-	-	-
1	-	-	-	-	-

Bing interpolated precision					
Standard recall	Query 1	Query 2	Query 3	Query 4	Average
0.1	0.6	0.8	1	1	0.85
0.2	0.5	0.47	0.9	1	0.72
0.3	0.4	0.45	0.75	0.47	0.52
0.4	0	0	0.72	0.4	0.28
0.5	-	-	-	-	-
0.6	-	-	-	-	-
0.7	-	-	-	-	-
0.8	-	-	-	-	-
0.9	-	-	-	-	-
1	-	-	-	-	-

DuckDuckGo interpolated precision					
Standard recall	Query 1	Query 2	Query 3	Query 4	Average
0.1	1	0.6	0.8	0.8	0.8
0.2	0.9	0.6	0.8	0.7	0.75
0.3	0.8	0.4	0.8	0.7	0.68
0.4	0.68	0.37	0.77	0.55	0.59
0.5	0	0	0.77	0.48	0.31
0.6	-	-	-	-	-
0.7	-	-	-	-	-
0.8	-	-	-	-	-
0.9	-	-	-	-	-
1	-	-	-	-	-



Conclusion:

Given the interpolated curves, Google would be considered the superior search engine. At any given recall-value, google would outperform Bing and DuckDuckGo in terms of precision (at least up to 60% recall) although it's quite plausible it would outperform for higher recall values as well. Although the judgement of the relevant links is subjective, one can argue in accordance with concordat's jury theorem (given that we are above 50% likely to judge each link correctly) that the overall results of the four of us should be "quite" accurate (although we did split up the work based on the query). However, we were quite novice in searching on other engines than google, but we found that Bing uses the same operators, but DuckDuckGo were a bit more limited. This could however have an impact on the results we got, and hence the performance.

Appendix:

Information need: “Any news about the war in Syria?”

Query 1: “News war+Syria”

The following 37 links was considered as relevant:

- <https://www.aljazeera.com/topics/country/syria.html> (Google, BING, DuckDuckGo)
- <https://www.newsnow.co.uk/h/World+News/War+&+Terrorism/Wars/War+in+Syria> (Google, BING, DuckDuckGo)
- <https://syria.liveuamap.com/> (Google, BING, DuckDuckGo)
- <https://www.bbc.com/news/world-middle-east-50250330> (Google)
- <https://www.independent.co.uk/topic/Syria> (Google, BING, DuckDuckGo)
- <https://www.reuters.com/places/syria> (Google, BING, DuckDuckGo)
- <https://www.theguardian.com/world/2019/oct/26/syria-turkey-arab-videos-torture-kurdish-bodies-militia> (Google)
- <https://www.theguardian.com/world/syria> (Google, BING, DuckDuckGo)
- <https://time.com/5711596/syria-war-crimes/> (Google)
- <https://www.nytimes.com/2019/10/14/world/middleeast/turkey-syria.html> (Google)
- <https://www.amnesty.org/en/latest/news/2019/10/syria-damning-evidence-of-war-crimes-and-other-violations-by-turkish-forces-and-their-allies/> (Google)
- https://news.yahoo.com/china-learning-russias-war-syria-183000614.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlMnNvbS8&guce_sig=AQAAAG9L2f-axah3iUJ3Q1zweocJZalow-luTHCalvCT7JJ1c9GOyls34NURZXPsvHusHVap9_-mEpFA2dH4QEnMnNCJEWQH4-hstHWVOUXPKnHY64ETNKGhp5TJ5WPUabVfF_DONHQo4Hj8YUqvywvWouC19z_NSBXcov7R0PQChSVG (Google)
- <https://news.un.org/en/story/2019/10/1050241> (Google)
- <https://news.sky.com/story/syria-sky-news-footage-almost-certainly-shows-russia-committed-war-crime-says-lawyer-11862341> (Google)
- <https://edition.cnn.com/specials/middleeast/syria> (Google, BING, DuckDuckGo)
- <https://abcnews.go.com/Politics/keeping-oil-syria-trump-considered-war-crime/story?id=66589757> (Google)
- <https://www.cbsnews.com/syria-crisis/> (Google)
- <https://www.cbsnews.com/news/national-security-adviser-robert-obrien-very-concerned-war-crimes-syria-turkish-offensive-face-the-nation/> (Google)
- <https://www.bbc.com/news/world-middle-east-17258397> (Google, BING, DuckDuckGo)
- <https://www.telegraph.co.uk/syria/> (Google, BING, DuckDuckGo)
- <https://www.foxnews.com/category/world/conflicts/syria> (Google)
- <https://www.nbcnews.com/syrias-suffering-families> (Google)
- <https://twitter.com/syriawar2> (Google, DuckDuckGo)
- <https://foreignpolicy.com/2019/10/28/putin-erdogan-deal-syria-kurds-agreement-war-continues/> (Google)
- <https://www.dailymail.co.uk/news/syria/index.html> (Google, BING)
- <https://www.vox.com/2019/10/21/20924757/syria-news-iraq-fruit-troops-trump> (Google)
- https://www.washingtonpost.com/gdpr-consent/?destination=%2fnational-security%2fallies-seek-reassurances-from-trump-administration-as-us-policy-on-syria-shifts%2f2019%2f11%2f15%2f58e71fba-07c0-11ea-818c-fcc65139e8c2_story.html%3f (Google)
- <https://www.almasdarnews.com/article/category/syria/> (DuckDuckGo, BING)
- <https://sana.sy/en/> (DuckDuckGo)
- <https://syriancivilwarmap.com/> (DuckDuckGo)
- <https://www.aljazeera.com/topics/subjects/turkey-syria-border.html> (DuckDuckGo)
- <https://www.syrianews.cc/> (DuckDuckGo)
- <https://www.nytimes.com/topic/destination/syria> (DuckDuckGo)
- <https://www.bbc.com/news/topics/cx1m7zg0w5zt/syria> (DuckDuckGo)
- <http://www.hurriyetdailynews.com/search/syrian%20war> (DuckDuckGo)
- <https://www.newsnow.co.uk/h/World+News/Middle+East/Syria> (BING)
- <https://www.newsdeeply.com/syria> (BING)

Information need: “You are burned out and you are looking for psychological advice for highly stressed people.”

Query 2: psychological advice + burned out + relieve stress

The following 26 links was considered as relevant:

- <https://www.helpguide.org/articles/stress/burnout-prevention-and-recovery.htm> (Google, Bing, DuckDuckGo)
- <https://www.psychologytoday.com/us/basics/burnout> (Google, Bing, DuckDuckGo)
- <https://qz.com/work/1640624/a-psychotherapist-explains-how-to-deal-with-burnout-symptoms/> (Google)
- <https://www.psychologytoday.com/us/blog/how-be-yourself/201611/how-recover-job-stress-and-burnout> (Google)
- <https://www.verywellmind.com/stress-and-burnout-symptoms-and-causes-3144516> (Google, Bing, DuckDuckGo)
- <https://blog.rescuetime.com/burnout-syndrome-recovery/> (Google)
- <https://www.mindtools.com/pages/article/recovering-from-burnout.htm> (Google, Bing, DuckDuckGo)
- <https://www.apa.org/monitor/2018/02/ce-corner> (Google)
- <https://www.thebalancecareers.com/how-to-deal-with-work-burnout-4142144> (Google)
- <https://advice.shinetext.com/articles/10-ways-to-bounce-back-from-burnout/> (Google)
- <https://www.medicalnewstoday.com/articles/325943.php#11> (Google)
- <https://psychcentral.com/blog/5-ways-to-cope-with-burnout/> (Google)
- <https://www.mentalhealth.org.uk/publications/how-manage-and-reduce-stress> (Google)
- <https://www.healthline.com/health/tips-for-identifying-and-preventing-burnout#prevention> (Google)
- <https://www.brisbanecitypsychologist.com.au/stress/> (Google)
- <https://www.mayoclinic.org/healthy-lifestyle/adult-health/in-depth/burnout/art-20046642> (Google)
- <https://www.cnet.com/how-to/signs-you-have-burnout/> (Google)
- <https://www.psychologytoday.com/us/blog/prescriptions-life/201907/4-things-do-when-youre-burning-out-work> (Bing, DuckDuckGo)
- <https://www.forbes.com/sites/travisbradberry/2016/11/08/7-powerful-ways-to-beat-burnout/#52b6879a61e6> (Bing, DuckDuckGo)
- <https://positivepsychology.com/stress-management-techniques-tips-burn-out/> (Bing, DuckDuckGo)
- <https://www.fashionbeans.com/article/burnout-symptoms-tips-recovery/> (Bing)
- <https://blogs.webmd.com/mental-health/20180116/5-signs-that-youre-burned-out> (Bing, DuckDuckGo)
- <https://joe.org/joe/1986spring/a1.php> (Bing)
- <https://www.mentalhelp.net/stress/management-techniques/> (DuckDuckGo)
- <https://www.healthline.com/health/signs-you-are-headed-for-a-burnout-at-work#1> (DuckDuckGo)
- <https://money.cnn.com/2018/04/18/pf/jobs/workplace-stress-burnout-signs/index.html> (DuckDuckGo)

Information need: “You want to find out how hashing and hash tables work.”

Query 3: “How hashing + hashtables work”

The following 45 links was considered as relevant:

- <https://www.hackerearth.com/practice/data-structures/hash-tables/basics-of-hash-tables/tutorial/> (Google, Bing, Duckduckgo)
- https://en.wikipedia.org/wiki/Hash_table (Google, Bing, Duckduckgo)
- <https://www.youtube.com/watch?v=shs0KM3wKv8> (Google, Bing)
- https://www.youtube.com/watch?v=KyUTuwz_b7Q (Google, Bing, Duckduckgo)
- <https://www.youtube.com/watch?v=MfhjkfocRR0> (Google, Duckduckgo)
- <https://yourbasic.org/algorithms/hash-tables-explained/> (Google)
- <https://stackoverflow.com/questions/730620/how-does-a-hash-table-work> (Google, Bing, Duckduckgo)
- <https://towardsdatascience.com/hash-tables-explained-5dc457db50da> (Google)
- <https://medium.com/@bartobri/hash-crash-the-basics-of-hash-tables-bef82a8ea550> (Google)
- <https://www.freecodecamp.org/forum/t/how-the-hash-table-data-structure-works-and-how-hashing-functions-work-a-tutorial-with-examples-in-python-and-java/19437> (Google)
- https://www.tutorialspoint.com/data_structures_algorithms/hash_data_structure.htm (Google, Bing)
- <https://www.geeksforgeeks.org/hashing-data-structure/> (Google)
- <https://www.geeksforgeeks.org/hashing-set-1-introduction/> (Google)
- <https://odino.org/this-is-how-a-dumb-hashtable-works/> (Google)
- https://everythingcomputerscience.com/discrete_mathematics/Data_Structures/Hash_Table.html (Google)
- <https://blog.cloudboost.io/data-structures-hash-tables-fd1cf3daba8> (Google)
- https://www.cs.auckland.ac.nz/software/AlgAnim/hash_tables.html (Google)
- <https://www.cs.cornell.edu/courses/cs3110/2014fa/lectures/13/lec13.html> (Google)
- <https://runestone.academy/runestone/books/published/pythonds/SortSearch/Hashing.html> (Google)
- <https://www.raywenderlich.com/206-swift-algorithm-club-hash-tables> (Google)
- <https://www.interviewcake.com/concept/java/hash-map> (Google)
- <https://howtodoinjava.com/java/collections/hashtable-class/> (Google)
- <https://www.sparknotes.com/cs/searching/hashtables/section1/> (Bing, Duckduckgo)
- <https://docs.microsoft.com/en-us/dotnet/api/system.collections.hashtable?view=netframework-4.8> (Bing, Duckduckgo)
- https://en.wikipedia.org/wiki/Hash_function (Bing, Duckduckgo)
- <https://www.interviewbit.com/tutorial/hash-tables/> (Bing, Duckduckgo)
- <http://www.cs.yale.edu/homes/aspnes/pinewiki/HashTables.html> (Bing, Duckduckgo)
- <https://www.mindprod.com/jgloss/hashtable.html> (Bing, Duckduckgo)
- <https://www.kindsonthegenius.com/java/2019/01/25/28-java-hashtables/> (Bing, Duckduckgo)
- <http://www.geocities.ws/rahuljg/Hashtables.htm> (Bing)
- <https://sites.google.com/site/usfcomputerscience/hashtables> (Bing)
- https://www.sas.com/content/dam/SAS/en_ca/User%20Group%20Presentations/Regina-User-Group/ShawnKaufmann-IntroHash-Sept2013.pdf (Bing, Duckduckgo)
- <https://www.cs.cmu.edu/~adamchik/15-121/lectures/Hashing/hashing.html> (Bing)
- <http://people.cs.aau.dk/~adavid/teaching/AA/AA1-06/08-Hashing+HashTables.pdf> (Bing)
- <https://www.c-sharpcorner.com/article/theory-of-hashing-and-hash-tables/> (Bing, Duckduckgo)
- <https://www.c-sharpcorner.com/blogs/hashtable-in-c-sharp-with-example> (Bing)
- <https://algs4.cs.princeton.edu/34hash/> (Bing, Duckduckgo)
- <https://stackoverflow.com/questions/38123760/how-does-hashing-work-for-python-sets> (Duckduckgo)
- <https://www.quora.com/How-do-hash-tables-work-How-do-distributed-hash-tables-work> (Duckduckgo)
- <https://courses.cs.northwestern.edu/311/html/hashtables.html> (Duckduckgo)
- <https://www.codeguru.com/columns/vb/working-with-hashtables-in-.net.htm> (Duckduckgo)
- <https://www.quora.com/What-are-hash-tables-and-why-are-they-useful> (Duckduckgo)
- <http://www.javacoffeebreak.com/faq/faq0034.html> (Duckduckgo)
- https://www.kirupa.com/net/dictionary_hashtable4.htm (Duckduckgo)
- https://www.cs.cornell.edu/courses/cs4120/2016sp/lectures/lec_hashtables/index.html (Duckduckgo)

Information need: “You are looking for information on the history of La Liga.”

Query 4: history of + la liga

The following 23 links was considered as relevant:

- https://el.wikipedia.org/wiki/%CE%A0%CF%81%CE%B9%CE%BC%CE%AD%CF%81%CE%B1_%CE%9D%CF%84%CE%B9%CE%B2%CE%B9%CF%83%CE%B9%CF%8C%CE%BD (google)
- https://en.wikipedia.org/wiki/La_Liga (google,Bing,duckduckgo)
- <https://www.footballhistory.org/league/la-liga.html> (google,Bing,duckduckgo)
- <https://www.liveabout.com/spanish-league-winners-and-runners-up-3557342> (google,Bing,duckduckgo)
- <https://www.espanol.com/a-history-of-la-liga-and-spanish-football/> (google, duckduckgo)
- <https://www.dummysports.com/la-liga-winners-list/> (google)
- <https://fbref.com/en/comps/12/history/La-Liga-Seasons> (google)
- <https://www.topendsports.com/sport/soccer/list-league-la-liga.htm> (google,bing,duckduckgo)
- <https://www.sportskeeda.com/football/history-of-total-goals-scored-by-teams-in-spanish-laliga-1> (google)
- <https://www.chaseyoursport.com/Football/6-Greatest-La-Liga-races-in-History/1188> (google)
- https://en.wikipedia.org/wiki/List_of_Spanish_football_champions (bing)
- <http://www.laligaweekly.com/history-of-la-liga> (bing)
- <https://www.skysports.com/la-liga-table> (bing)
- https://sco.wikipedia.org/wiki/La_Liga (bing)
- <https://www.casinouk.com/sports-betting/football-betting/la-liga-league> (bing)
- https://en.wikipedia.org/wiki/List_of_Spanish_football_champions (duckduckgo)
- www.laligaweekly.com/history-of-la-liga (duckduckgo)
- <https://www.soccertimes.com/leagues/spanish-la-liga> (duckduckgo)
- https://en.wikipedia.org/wiki/La_Liga_Awards (duckduckgo)
- <https://dirtyfilthyandmessi.wordpress.com/history-of-la-liga/> (duckduckgo)
- <http://www.instantactionsports.com/sportsbook/soccer/spanish-soccer-league-history.html> (duckduckgo)
- <https://lavinlaliga.blogspot.com/p/history-of-la-liga.html> (duckduckgo)
- <http://foxsportsstories.com/2017/10/25/the-history-of-spanish-la-liga-1929-2017/> (duckduckgo)