

```

python3 protein_search.py \
-d protein_vectors.fvecs \
-q targets.fasta \
-s swissprot_50k.fasta \
-o benchmarks/res_lsh_k6_L8_w1.txt \
-method lsh \
-N 50 \
--lsh_k 6 \
--lsh_w 1.0

```

Query Protein: A0A009I3Y5

N = 50 (μέγεθος λίστας Top-N για την αξιολόγηση Recall@N)

[1] Συνοπτική σύγκριση μεθόδων

Method	Time/query(s)	QPS	Recall@N vs BLAST
Euclidean LSH	0.248	4	0.04
BLAST (Ref)	0.865	1	1.00 (ref)

[2] Top-N γείτονες ανά μέθοδο (εδώ N = 10 για εκτύπωση)

Method: Euclidean LSH

Rank	Neighbor ID	L2 Dist	BLAST Identity	In BLAST Top-N?	Bio comment
1	sp A5UG81 KDKA_HAEIG	0.35	1%	No	Possible false positive
2	sp P44033 Y665_HAEIN	0.36	1%	No	Possible false positive
3	sp P14181 LICA2_HAEIF	0.36	2%	No	Possible false positive
4	sp Q8EKA0 SELU_SHEON	0.39	3%	No	Possible false positive

5	sp B2K7I7 RAVA_YERPB	0.39	2%	No	Possible false positive
6	sp Q12SQ2 SELU_SHEDO	0.40	3%	No	Possible false positive
7	sp B4F0D4 VIAA_PROMH	0.41	1%	No	Possible false positive
8	sp A1SW96 MNMC_PSYIN	0.41	0%	No	Possible false positive
9	sp A3DA85 SELU_SHEB5	0.41	3%	No	Possible false positive
10	sp Q4QLU9 RLMC_HAEI8	0.41	0%	No	Possible false positive

Query Protein: A0A009PCK4

N = 50 (μέγεθος λίστας Top-N για την αξιολόγηση Recall@N)

[1] Συνοπτική σύγκριση μεθόδων

Method	Time/query(s)	QPS	Recall@N vs BLAST
Euclidean LSH	0.248	4	0.04
BLAST (Ref)	0.346	2	1.00 (ref)

[2] Top-N γείτονες ανά μέθοδο (εδώ N = 10 για εκτύπωση)

Method: Euclidean LSH

Rank	Neighbor ID	L2 Dist	BLAST Identity	In BLAST Top-N?	Bio comment
1	sp A5UG81 KDKA_HAEIG	0.34	1%	No	Possible false positive
2	sp P44033 Y665_HAEIN	0.36	1%	No	Possible false positive
3	sp P14181 LICA2_HAEIF	0.36	2%	No	Possible false positive
4	sp Q8EKA0 SELU_SHEON	0.39	3%	No	Possible false positive
5	sp B2K7I7 RAVA_YERPB	0.40	2%	No	Possible false positive

6	sp Q12SQ2 SELU_SHEDO	0.40	3%	No	Possible false positive
7	sp A1SW96 MNMC_PSYIN	0.41	0%	No	Possible false positive
8	sp B4F0D4 VIAA_PROMH	0.41	1%	No	Possible false positive
9	sp A3DA85 SELU_SHEB5	0.41	3%	No	Possible false positive
10	sp Q4QLU9 RLMC_HAEI8	0.41	0%	No	Possible false positive

Query Protein: A0A001

N = 50 (μέγεθος λίστας Top-N για την αξιολόγηση Recall@N)

[1] Συνοπτική σύγκριση μεθόδων

Method	Time/query(s)	QPS	Recall@N vs BLAST
Euclidean LSH	0.248	4	0.20
BLAST (Ref)	0.734	1	1.00 (ref)

[2] Top-N γείτονες ανά μέθοδο (εδώ N = 10 για εκτύπωση)

Method: Euclidean LSH

Rank	Neighbor ID	L2 Dist	BLAST Identity	In BLAST Top-N?	Bio comment
1	sp P9WQJ3 FATRP_MYCTU	0.30	1%	Yes	-
2	sp Q13BH6 NDVA_RHOPS	0.32	1%	Yes	-
3	sp Q03024 APRD_PSEAE	0.33	1%	Yes	-
4	sp Q20Z38 NDVA_RHOPB	0.34	1%	Yes	-
5	sp Q9ZEP3 CSEC_STRCO	0.39	1%	No	Possible false positive
6	sp A0K739 SSUB1_BURCH	0.40	2%	No	Possible false positive

7	sp Q51719 YCOBA_PROFR	0.40	1%	No	Possible false positive
8	sp Q8P8W4 MSBA_XANCP	0.40	1%	Yes	-
9	sp Q3JSR6 SSUB_BURP1	0.41	2%	No	Possible false positive
10	sp P23886 CYDC_ECOLI	0.41	1%	Yes	-

Query Protein: A0A002

N = 50 (μέγεθος λίστας Top-N για την αξιολόγηση Recall@N)

[1] Συνοπτική σύγκριση μεθόδων

Method	Time/query(s)	QPS	Recall@N vs BLAST
Euclidean LSH	0.248	4	0.24
BLAST (Ref)	0.649	1	1.00 (ref)

[2] Top-N γείτονες ανά μέθοδο (εδώ N = 10 για εκτύπωση)

Method: Euclidean LSH

Rank	Neighbor ID	L2 Dist	BLAST Identity	In BLAST Top-N?	Bio comment
1	sp P9WQJ3 FATRP_MYCTU	0.31	1%	Yes	-
2	sp Q13BH6 NDVA_RHOPS	0.33	2%	Yes	-
3	sp Q03024 APRD_PSEAE	0.35	0%	Yes	-
4	sp Q20Z38 NDVA_RHOPB	0.36	2%	Yes	-
5	sp P23886 CYDC_ECOLI	0.41	1%	Yes	-
6	sp Q8P8W4 MSBA_XANCP	0.41	1%	Yes	-
7	sp Q51719 YCOBA_PROFR	0.42	1%	No	Possible false positive

8	sp Q82MV1 SSUB1_STRAW	0.42	1%	No	Possible false positive
9	sp Q28433 TAP1_GORGO	0.42	0%	Yes	-
10	sp A0K739 SSUB1_BURCH	0.43	1%	No	Possible false positive

Query Protein: A0A009HL96

N = 50 (μέγεθος λίστας Top-N για την αξιολόγηση Recall@N)

[1] Συνοπτική σύγκριση μεθόδων

Method	Time/query(s)	QPS	Recall@N vs BLAST
Euclidean LSH	0.248	4	0.66
BLAST (Ref)	0.276	3	1.00 (ref)

[2] Top-N γείτονες ανά μέθοδο (εδώ N = 10 για εκτύπωση)

Method: Euclidean LSH

Rank	Neighbor ID	L2 Dist	BLAST Identity	In BLAST Top-N?	Bio comment
1	sp Q5HI09 GRAR_STAAC	0.18	1%	Yes	-
2	sp P33112 SPAR_BACIU	0.18	2%	Yes	-
3	sp Q99VW2 GRAR_STAAN	0.19	1%	Yes	-
4	sp Q8CQ37 GRAR_STAES	0.19	2%	Yes	-
5	sp Q99RR6 HSSR_STAAM	0.21	1%	Yes	-
6	sp P28257 YCF27_GALSU	0.22	2%	Yes	-
7	sp Q9AE24 RPRY_BACFR	0.22	2%	Yes	-
8	sp P0CAW8 CTRA_CAUVC	0.22	1%	Yes	-

9	sp A6WZ81 CTRA_BRUA4	0.24	1%	Yes	-
10	sp P45189 PHOB_HAEIN	0.25	1%	Yes	-

Query Protein: A0A009HLV9

N = 50 (μέγεθος λίστας Top-N για την αξιολόγηση Recall@N)

[1] Συνοπτική σύγκριση μεθόδων

Method	Time/query(s)	QPS	Recall@N vs BLAST
--------	---------------	-----	-------------------

Euclidean LSH	0.248	4	0.36
---------------	-------	---	------

BLAST (Ref)	0.452	2	1.00 (ref)
-------------	-------	---	------------

[2] Top-N γείτονες ανά μέθοδο (εδώ N = 10 για εκτύπωση)

Method: Euclidean LSH

Rank	Neighbor ID	L2 Dist	BLAST Identity	In BLAST Top-N?	Bio comment
------	-------------	---------	----------------	-----------------	-------------

1	sp P28787 NTRC_PROHU	0.14	1%	Yes	-
2	sp O87455 LUXO_VIBC3	0.22	3%	Yes	-
3	sp Q06065 ATOC_ECOLI	0.22	1%	Yes	-
4	sp Q5E3W8 NORR_ALIF1	0.22	3%	Yes	-
5	sp P36219 WTSA_PANSE	0.22	0%	Yes	-
6	sp Q3K999 SFNR_PSEPF	0.23	2%	Yes	-
7	sp A1KU52 RUVB_NEIMF	0.24	0%	No	Possible false positive
8	sp P23747 ALGB_PSEAE	0.24	1%	Yes	-
9	sp P59402 NORR_SHIFL	0.24	1%	Yes	-

10	sp P37931 HRPS_PSESY	0.24	1%	Yes	-
----	----------------------	------	----	-----	---

Query Protein: A0A009HN45

N = 50 (μέγεθος λίστας Τop-N για την αξιολόγηση Recall@N)

[1] Συνοπτική σύγκριση μεθόδων

Method	Time/query(s)	QPS	Recall@N vs BLAST
Euclidean LSH	0.248	4	0.24
BLAST (Ref)	0.656	1	1.00 (ref)

[2] Top-N γείτονες ανά μέθοδο (εδώ N = 10 για εκτύπωση)

Method: Euclidean LSH

Rank	Neighbor ID	L2 Dist	BLAST Identity	In BLAST Top-N?	Bio comment
1	sp A6QKD8 SECA2_STAAE	0.14	1%	No	Possible false positive
2	sp A7MSB2 RAPA_VIBC1	0.14	1%	Yes	-
3	sp Q3K063 SECA2_STRA1	0.15	1%	No	Possible false positive
4	sp Q9ZDW2 UVRB_RICPR	0.15	1%	No	Possible false positive
5	sp Q1MQP3 SECA_LAWIP	0.15	1%	No	Possible false positive
6	sp Q7VJC6 SECA_HELHP	0.16	1%	No	Possible false positive
7	sp B0BT63 RAPA_ACTPJ	0.16	1%	Yes	-
8	sp Q92H92 SECA_RICCN	0.16	1%	No	Possible false positive
9	sp P52126 ABPB_ECOLI	0.16	1%	No	Possible false positive
10	sp Q0HMR2 RAPA_SHESM	0.16	1%	Yes	-

Query Protein: A0A009HQC9

N = 50 (μέγεθος λίστας Top-N για την αξιολόγηση Recall@N)

[1] Συνοπτική σύγκριση μεθόδων

Method	Time/query(s)	QPS	Recall@N vs BLAST
Euclidean LSH	0.248	4	0.26
BLAST (Ref)	0.675	1	1.00 (ref)

[2] Top-N γείτονες ανά μέθοδο (εδώ N = 10 για εκτύπωση)

Method: Euclidean LSH

Rank	Neighbor ID	L2 Dist	BLAST Identity	In BLAST Top-N?	Bio comment
1	sp Q0HR19 RAPA_SHESR	0.11	1%	Yes	-
2	sp Q0HMR2 RAPA_SHESM	0.11	1%	Yes	-
3	sp A0KGL5 RAPA_AERHH	0.11	1%	Yes	-
4	sp A1A7A6 RAPA_ECOK1	0.12	1%	Yes	-
5	sp A7ZW09 RAPA_ECOHS	0.12	1%	Yes	-
6	sp B0BT63 RAPA_ACTPJ	0.12	1%	Yes	-
7	sp A7MSB2 RAPA_VIBC1	0.12	1%	Yes	-
8	sp Q48LP5 RAPA_PSE14	0.13	2%	Yes	-
9	sp B5FI41 RAPA_SALDC	0.13	1%	Yes	-
10	sp Q5PDF0 RAPA_SALPA	0.13	1%	Yes	-

Query Protein: A0A009HPM0

N = 50 (μέγεθος λίστας Top-N για την αξιολόγηση Recall@N)

[1] Συνοπτική σύγκριση μεθόδων

Method	Time/query(s)	QPS	Recall@N vs BLAST
Euclidean LSH	0.248	4	0.06
BLAST (Ref)	0.535	1	1.00 (ref)

[2] Top-N γείτονες ανά μέθοδο (εδώ N = 10 για εκτύπωση)

Method: Euclidean LSH

Rank	Neighbor ID	L2 Dist	BLAST Identity	In BLAST Top-N?	Bio comment
1	sp Q612F5 PCCA_CAEBR	0.27	1%	Yes	-
2	sp Q5I0C3 MCCA_RAT	0.28	3%	Yes	-
3	sp Q9CMJ8 RIMK_PASMU	0.29	1%	No	Possible false positive
4	sp Q12KL2 G6PI_SHEDO	0.29	0%	No	Possible false positive
5	sp P44440 RNB_HAEIN	0.30	0%	No	Possible false positive
6	sp Q9CP74 RNB_PASMU	0.31	0%	No	Possible false positive
7	sp Q8DXM9 GSHAB_STRAS	0.31	1%	No	Possible false positive
8	sp A5F3L5 MUTL_VIBC3	0.31	1%	No	Possible false positive
9	sp P57029 DNAG_NEIMB	0.31	1%	No	Possible false positive
10	sp P07244 PUR2_YEAST	0.31	1%	No	Possible false positive

Query Protein: A0A009IB02

N = 50 (μέγεθος λίστας Top-N για την αξιολόγηση Recall@N)

[1] Συνοπτική σύγκριση μεθόδων

Method	Time/query(s)	QPS	Recall@N vs BLAST
Euclidean LSH	0.248	4	0.29
BLAST (Ref)	0.808	1	1.00 (ref)

[2] Top-N γείτονες ανά μέθοδο (εδώ N = 10 για εκτύπωση)

Method: Euclidean LSH

Rank	Neighbor ID	L2 Dist	BLAST Identity	In BLAST Top-N?	Bio comment
1	sp P10802 ODP2_AZ0VI	0.28	1%	Yes	-
2	sp P11961 ODP2_GEOSE	0.46	1%	Yes	-
3	sp Q9SQI8 ODP24_ARATH	0.46	1%	Yes	-
4	sp P49786 BCCP_BACSU	0.46	15%	No	-
5	sp P22439 ODPX_BOVIN	0.50	1%	Yes	-
6	sp P43874 BCCP_HAEIN	0.51	16%	Yes	-
7	sp P29337 BCCP_STRMU	0.53	11%	Yes	-
8	sp P02904 BCCP_PROFR	0.55	18%	Yes	-
9	sp Q06881 BCCP_NOSS1	0.56	12%	Yes	-
10	sp P0ABE1 BCCP_SHIFL	0.56	16%	Yes	-

Query Protein: A0A010Q3W2

N = 50 (μέγεθος λίστας Top-N για την αξιολόγηση Recall@N)

[1] Συνοπτική σύγκριση μεθόδων

Method	Time/query(s)	QPS	Recall@N vs BLAST
Euclidean LSH	0.248	4	0.11
BLAST (Ref)	0.541	1	1.00 (ref)

[2] Top-N γείτονες ανά μέθοδο (εδώ N = 10 για εκτύπωση)

Method: Euclidean LSH

Rank	Neighbor ID	L2 Dist	BLAST Identity	In BLAST Top-N?	Bio comment
1	sp D1ZEM6 LIS12_SORMK	0.24	2%	No	Possible false positive
2	sp P0CS56 CMR1_CRYNJ	0.25	1%	No	Possible false positive
3	sp Q2UUT4 CMR1_ASPOR	0.25	1%	No	Possible false positive
4	sp P0CS29 ATG18_CRYNB	0.26	1%	No	Possible false positive
5	sp A2QI22 YTM1_ASPNC	0.27	1%	No	Possible false positive
6	sp A1CXL0 YTM1_NEOFI	0.27	1%	No	Possible false positive
7	sp A2RAG5 ATG18_ASPNC	0.28	1%	No	Possible false positive
8	sp O35828 CORO7_RAT	0.28	1%	No	Possible false positive
9	sp C1GB49 LIS1_PARBD	0.28	1%	No	Possible false positive
10	sp Q9P7C0 YKU5_SCHPO	0.29	0%	No	Possible false positive

Query Protein: A0A010Q8R4

N = 50 (μέγεθος λίστας Top-N για την αξιολόγηση Recall@N)

[1] Συνοπτική σύγκριση μεθόδων

Method	Time/query(s)	QPS	Recall@N vs BLAST
Euclidean LSH	0.248	4	0.10
BLAST (Ref)	0.668	1	1.00 (ref)

[2] Top-N γείτονες ανά μέθοδο (εδώ N = 10 για εκτύπωση)

Method: Euclidean LSH

Rank	Neighbor ID	L2 Dist	BLAST Identity	In BLAST Top-N?	Bio comment
1	sp Q9P4R5 CREC_EMENI	0.25	0%	No	Possible false positive
2	sp Q652L2 HIRA_ORYSJ	0.25	0%	No	Possible false positive
3	sp P0CS56 CMR1_CRYNJ	0.25	1%	No	Possible false positive
4	sp P0CS29 ATG18_CRYNB	0.26	0%	No	Possible false positive
5	sp A1DIN7 ATP25_NEOFI	0.27	0%	No	Possible false positive
6	sp Q5NVM2 DC1I2_PONAB	0.28	1%	No	Possible false positive
7	sp A2QGF0 NAGS_ASPNC	0.28	1%	No	Possible false positive
8	sp Q6CXX3 HIR1_KLULA	0.28	0%	Yes	-
9	sp Q2UUT4 CMR1_ASPOR	0.28	0%	No	Possible false positive
10	sp C3VEQ3 NCED_ONCHC	0.28	1%	No	Possible false positive