

```
python3 protein_nlsh_build.py \
-d protein_vectors.fvecs \
-i protein_index \
-m 15 \
--knn 40 \
--epochs 70 \
--lr 0.00015
```

```
=====
=====
```

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	

BLAST(Ref)	0.000	0	1.00	(Reference)
neural	0.4692	2	0.000	

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

Method: neural

Rank	Neighbor ID	L2Dist	Ident.	In BLAST?	Bio Comment

1	sp Q4QLU9 RLMC_HAEI8	0.4106	0.3%	No	RLMC_HAEI8 [Possible Remote Homolog?]
2	sp P24362 KRB2_VACCW	0.4235	0.4%	No	KRB2_VACCW [Possible Remote Homolog?]

3	sp Q8EIL1 RSMC_SHEON 0.4235	1.1%	No	RSMC_SHEON [Possible Remote Homolog?]
4	sp B9IWP6 MTNK_BACCQ 0.4284	1.3%	No	MTKN_BACCQ [Possible Remote Homolog?]
5	sp Q7VP87 RSMC_HAEDU 0.4288	1.4%	No	RSMC_HAEDU [Possible Remote Homolog?]
6	sp Q7MHY6 RSMC_VIBVY 0.4293	1.4%	No	RSMC_VIBVY [Possible Remote Homolog?]
7	sp A9KYL6 UBIB_SHEB9 0.4311	1.0%	No	UBIB_SHEB9 [Possible Remote Homolog?]
8	sp Q6C2A3 BUD32_Y... 0.4327	2.5%	No	BUD32_YARLI [Possible Remote Homolog?]
9	sp Q8E9R5 UBIB_SHEON 0.4365	1.0%	No	UBIB_SHEON [Possible Remote Homolog?]
10	sp B1KD54 RLMG_SHEWM 0.4380	2.2%	No	RLMG_SHEWM [Possible Remote Homolog?]

=====

=====

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	

BLAST(Ref)	1.387	0	1.00	(Reference)
neural	0.4692	2	0.041	

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

Method: neural

Rank	Neighbor ID	L2Dist	Ident.	In BLAST?	Bio Comment
1	sp Q4QLU9 RLMC_HAEI8	0.4115	0.3%	No	RLMC_HAEI8 [Possible Remote Homolog?]
2	sp P24362 KRB2_VACCW	0.4178	0.4%	Yes	KRB2_VACCW
3	sp B9IWP6 MTNK_BACCQ	0.4239	1.3%	No	MTNK_BACCQ [Possible Remote Homolog?]
4	sp Q8EIL1 RSMC_SHEON	0.4268	1.1%	No	RSMC_SHEON [Possible Remote Homolog?]
5	sp Q6C2A3 BUD32_Y...	0.4284	2.5%	No	BUD32_YARLI [Possible Remote Homolog?]
6	sp A9KYL6 UBIB_SHEB9	0.4298	1.0%	No	UBIB_SHEB9 [Possible Remote Homolog?]
7	sp Q7MHY6 RSMC_VIBVY	0.4298	1.4%	No	RSMC_VIBVY [Possible Remote Homolog?]
8	sp Q7VP87 RSMC_HAEDU	0.4303	1.4%	No	RSMC_HAEDU [Possible Remote Homolog?]
9	sp B0BBT2 PKND_CHLTB	0.4340	0.2%	Yes	PKND_CHLTB
10	sp Q8E9R5 UBIB_SHEON	0.4348	1.0%	No	UBIB_SHEON [Possible Remote Homolog?]

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
BLAST(Ref)	0.804	1	1.00 (Reference)
neural	0.4692	2	0.200

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

Method: neural

Rank	Neighbor ID	L2Dist	Ident.	In BLAST?	Bio Comment
------	-------------	--------	--------	-----------	-------------

1	sp P9WQJ3 FATRP_M...	0.2992	1.0%	Yes	FATRP_MYCTU
2	sp Q13BH6 NDVA_RHOPS	0.3194	1.2%	Yes	NDVA_RHOPS
3	sp Q03024 APRD_PSEAE	0.3297	0.7%	Yes	APRD_PSEAE
4	sp Q20Z38 NDVA_RHOPB	0.3435	1.2%	Yes	NDVA_RHOPB
5	sp A0K739 SSUB1_B...	0.3995	1.5%	No	SSUB1_BURCH [Possible Remote Homolog?]
6	sp Q51719 YCOBA_P...	0.4010	0.9%	No	YCOBA_PROFR [Possible Remote Homolog?]
7	sp Q8P8W4 MSBA_XANCP	0.4015	1.2%	Yes	MSBA_XANCP
8	sp Q3JSR6 SSUB_BURP1	0.4070	1.9%	No	SSUB_BURP1 [Possible Remote Homolog?]
9	sp P23886 CYDC_ECOLI	0.4095	0.7%	Yes	CYDC_ECOLI
10	sp Q21BF6 MODC_RHOPB	0.4150	0.6%	No	MODC_RHOPB [Possible Remote Homolog?]

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
--------	---------------	-----	-------------------

BLAST(Ref)	0.672	1	1.00	(Reference)
neural	0.4692	2	0.300	

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

Method: neural

Rank	Neighbor ID	L2Dist	Ident.	In BLAST?	Bio Comment
1	sp P9WQJ3 FATRP_M...	0.3105	1.2%	Yes	FATRP_MYCTU
2	sp Q13BH6 NDVA_RHOPS	0.3339	1.5%	Yes	NDVA_RHOPS
3	sp Q03024 APRD_PSEAE	0.3467	0.3%	Yes	APRD_PSEAE
4	sp Q20Z38 NDVA_RHOPB	0.3572	1.7%	Yes	NDVA_RHOPB
5	sp P23886 CYDC_ECOLI	0.4084	0.5%	Yes	CYDC_ECOLI
6	sp Q8P8W4 MSBA_XANCP	0.4088	0.7%	Yes	MSBA_XANCP
7	sp Q51719 YCOBA_P...	0.4176	0.9%	No	YCOBA_PROFR [Possible Remote Homolog?]
8	sp Q82MV1 SSUB1_S...	0.4216	0.7%	No	SSUB1_STRAW [Possible Remote Homolog?]
9	sp Q28433 TAP1_GORGO	0.4220	0.2%	Yes	TAP1_GORGO
10	sp A0K739 SSUB1_B...	0.4278	1.1%	No	SSUB1_BURCH [Possible Remote Homolog?]

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
--------	---------------	-----	-------------------

BLAST(Ref)	0.318	3	1.00	(Reference)
neural	0.4692	2	0.000	

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

Method: neural

Rank	Neighbor ID	L2Dist	Ident.	In BLAST?	Bio Comment

1	sp Q9ZWS7 ARR7_ARATH	0.4956	2.6%	No	ARR7_ARATH [Possible Remote Homolog?]
2	sp Q81IN8 METN2_B...	0.5379	1.0%	No	METN2_BACCR [Possible Remote Homolog?]
3	sp Q5HQE7 UGTP_STAEQ	0.5385	0.6%	No	UGTP_STAEQ [Possible Remote Homolog?]
4	sp Q13237 KGP2_HUMAN	0.5425	0.2%	No	KGP2_HUMAN [Possible Remote Homolog?]
5	sp B7HU46 UGTP_BACC7	0.5438	0.6%	No	UGTP_BACC7 [Possible Remote Homolog?]
6	sp B3LKQ3 GPI3_YEAS1	0.5446	0.3%	No	GPI3_YEAS1 [Possible Remote Homolog?]
7	sp Q6G828 PCRA_STAAS	0.5449	1.0%	No	PCRA_STAAS [Possible Remote Homolog?]
8	sp Q9KTJ5 METN_VIBCH	0.5459	1.0%	No	METN_VIBCH [Possible Remote Homolog?]
9	sp O61267 LOK_DROME	0.5480	0.6%	No	LOK_DROME [Possible Remote Homolog?]
10	sp Q8ELR4 POTA_OCEIH	0.5483	1.0%	No	POTA_OCEIH [Possible Remote Homolog?]

=====

=====

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
--------	---------------	-----	-------------------

BLAST(Ref)	0.490	2	1.00 (Reference)
------------	-------	---	------------------

neural	0.4692	2	0.000
--------	--------	---	-------

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

Method: neural

Rank	Neighbor ID	L2Dist	Ident.	In BLAST?	Bio Comment
------	-------------	--------	--------	-----------	-------------

1	sp P0A9K4 PHOL_ECOL6	0.3822	0.5%	No	PHOL_ECOL6 [Possible Remote Homolog?]
2	sp P0A9K5 PHOL_ECO57	0.3822	0.5%	No	PHOL_ECO57 [Possible Remote Homolog?]
3	sp Q6F9A8 CYSA_ACIAD	0.3913	1.9%	No	CYSA_ACIAD [Possible Remote Homolog?]
4	sp Q8ZQN7 DING_SALTY	0.3922	0.2%	No	DING_SALTY [Possible Remote Homolog?]
5	sp P56255 PCRA_GEOSE	0.3928	1.7%	No	PCRA_GEOSE [Possible Remote Homolog?]
6	sp B0BT63 RAPA_ACTPJ	0.3939	0.3%	No	RAPA_ACTPJ [Possible Remote Homolog?]
7	sp Q47EI2 UVRB_DECAR	0.3944	0.3%	No	UVRB_DECAR [Possible Remote Homolog?]
8	sp A7MSB2 RAPA_VIBC1	0.3972	1.5%	No	RAPA_VIBC1 [Possible Remote Homolog?]

9	sp Q9KLL9 MODC_VIBCH	0.3973	0.5%	No	MODC_VIBCH [Possible Remote Homolog?]
10	sp C5BEJ2 UVRB_EDWI9	0.3976	0.3%	No	UVRB_EDWI9 [Possible Remote Homolog?]

=====

=====

Query Protein: A0A009HN45

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

[1] Συνοπτική σύγκριση μεθόδων

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

BLAST(Ref)	0.699	1	1.00	(Reference)
------------	-------	---	------	-------------

neural	0.4692	2	0.235
--------	--------	---	-------

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

Method: neural

Rank	Neighbor ID	L2Dist	Ident.	In BLAST?	Bio Comment
------	-------------	--------	--------	-----------	-------------

1	sp A6QKD8 SECA2_S...	0.1428	0.7%	No	SECA2_STAAE [Possible Remote Homolog?]
---	----------------------	--------	------	----	--

2	sp A7MSB2 RAPA_VIBC1	0.1438	1.2%	Yes	RAPA_VIBC1
---	----------------------	--------	------	-----	------------

3	sp Q3K063 SECA2_S...	0.1516	0.6%	No	SECA2_STRA1 [Possible Remote Homolog?]
---	----------------------	--------	------	----	--

4	sp Q9ZDW2 UVRB_RICPR	0.1528	1.0%	No	UVRB_RICPR [Possible Remote Homolog?]
---	----------------------	--------	------	----	---------------------------------------

5	sp Q1MQP3 SECA_LAWIP	0.1536	0.7%	No	SECA_LAWIP [Possible Remote Homolog?]
---	----------------------	--------	------	----	---------------------------------------

6	sp Q7VJC6 SECA_HELHP	0.1598	0.9%	No	SECA_HELHP [Possible Remote Homolog?]
7	sp B0BT63 RAPA_ACTPJ	0.1604	0.5%	Yes	RAPA_ACTPJ
8	sp Q92H92 SECA_RICCN	0.1627	1.1%	No	SECA_RICCN [Possible Remote Homolog?]
9	sp P52126 ABPB_ECOLI	0.1632	1.1%	No	ABPB_ECOLI [Possible Remote Homolog?]
10	sp Q0HMR2 RAPA_SHESM	0.1647	0.7%	Yes	RAPA_SHESM

=====

=====

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
BLAST(Ref)	0.698	1	1.00 (Reference)
neural	0.4692	2	0.261

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

Method: neural

Rank	Neighbor ID	L2Dist	Ident.	In BLAST?	Bio Comment
1	sp Q0HR19 RAPA_SHESR	0.1070	0.8%	Yes	RAPA_SHESR
2	sp Q0HMR2 RAPA_SHESM	0.1075	0.8%	Yes	RAPA_SHESM
3	sp A0KGL5 RAPA_AERHH	0.1091	1.3%	Yes	RAPA_AERHH
4	sp A1A7A6 RAPA_ECOK1	0.1209	0.8%	Yes	RAPA_ECOK1

5	sp A7ZW09 RAPA_ECOHS 0.1221 0.8% Yes RAPA_ECOHS
6	sp B0BT63 RAPA_ACTPJ 0.1222 0.5% Yes RAPA_ACTPJ
7	sp A7MSB2 RAPA_VIBC1 0.1236 1.3% Yes RAPA_VIBC1
8	sp Q48LP5 RAPA_PSE14 0.1277 1.9% Yes RAPA_PSE14
9	sp B5FI41 RAPA_SALDC 0.1283 0.9% Yes RAPA_SALDC
10	sp Q5PDF0 RAPA_SALPA 0.1288 0.9% Yes RAPA_SALPA

=====

=====

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
BLAST(Ref)	0.520	1	1.00 (Reference)
neural	0.4692	2	0.062

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

Method: neural

Rank	Neighbor ID	L2Dist	Ident.	In BLAST?	Bio Comment
1	sp Q612F5 PCCA_CAEBR 0.2744 0.6% Yes PCCA_CAEBR				
2	sp Q5I0C3 MCCA_RAT 0.2814 3.0% Yes MCCA_RAT				
3	sp Q9CMJ8 RIMK_PASMU 0.2928 0.6% No RIMK_PASMU [Possible Remote Homolog?]				

4	sp Q12KL2 G6PI_SHEDO 0.2934 0.2% No G6PI_SHEDO [Possible Remote Homolog?]
5	sp B1KD54 RLMG_SHEWM 0.3169 1.1% No RLMG_SHEWM [Possible Remote Homolog?]
6	sp B2TWG0 TRMA_SHIB3 0.3169 0.8% No TRMA_SHIB3 [Possible Remote Homolog?]
7	sp Q6FAM6 PURT_ACIAD 0.3172 1.1% No PURT_ACIAD [Possible Remote Homolog?]
8	sp B8F632 G6PI_GLAP5 0.3173 0.7% No G6PI_GLAP5 [Possible Remote Homolog?]
9	sp B1LNS5 TRMA_ECOSM 0.3198 0.8% No TRMA_ECOSM [Possible Remote Homolog?]
10	sp Q88XP4 Y1151_L... 0.3204 0.4% No Y1151_LACPL [Possible Remote Homolog?]

=====

=====

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
BLAST(Ref)	0.806	1	1.00 (Reference)
neural	0.4692	2	0.029

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

Method: neural

Rank	Neighbor ID	L2Dist	Ident.	In BLAST?	Bio Comment
------	-------------	--------	--------	-----------	-------------

1	sp Q98D88 YIDC_RHILO 0.6544 0.2% No YIDC_RHILO
2	sp B0T274 MSCL_CAUSK 0.6596 12.7% No MSCL_CAUSK
3	sp Q7SFC8 ROK1_NEUCR 0.6649 0.8% No ROK1_NEUCR
4	sp Q4IJH1 DBP3_GIBZE 0.6651 0.8% No DBP3_GIBZE
5	sp A3N477 YIDC_BURP6 0.6653 0.2% No YIDC_BURP6
6	sp B1K0Y4 YIDC_BURO0 0.6653 0.2% No YIDC_BURO0
7	sp A1V7D5 YIDC_BURMS 0.6657 0.2% No YIDC_BURMS
8	sp B4E7D5 YIDC_BURCJ 0.6660 0.2% No YIDC_BURCJ
9	sp C3MF49 YIDC_SINFN 0.6670 1.1% No YIDC_SINFN
10	sp P65622 YIDC_BORPE 0.6679 0.2% No YIDC_BORPE

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
--------	---------------	-----	-------------------

BLAST(Ref)	0.664	1	1.00 (Reference)
------------	-------	---	------------------

neural	0.4692	2	0.000
--------	--------	---	-------

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

Method: neural

Rank	Neighbor ID	L2Dist	Ident.	In BLAST?	Bio Comment
------	-------------	--------	--------	-----------	-------------

```

-----
1 | sp|Q09499|IRE1_CAEEL | 0.3583 | 0.1% | No | IRE1_CAEEL [Possible Remote
Homolog?]
2 | sp|Q9EQY0|ERN1_MOUSE | 0.3701 | 0.1% | No | ERN1_MOUSE [Possible
Remote Homolog?]
3 | sp|Q5BCU8|ATG1_EMENI | 0.3740 | 1.0% | No | ATG1_EMENI [Possible
Remote Homolog?]
4 | sp|Q7XIT1|IRE1_ORYSJ | 0.3748 | 1.3% | No | IRE1_ORYSJ [Possible Remote
Homolog?]
5 | sp|Q4UWF4|XOPAC_X... | 0.3775 | 0.9% | No | XOPAC_XANC8 [Possible
Remote Homolog?]
6 | sp|Q7RX99|ATG1_NEUCR | 0.3876 | 0.8% | No | ATG1_NEUCR [Possible
Remote Homolog?]
7 | sp|Q2PAY2|ROP18_T... | 0.3889 | 0.2% | No | ROP18_TOXGO [Possible
Remote Homolog?]
8 | sp|A0A0P0VIP0|LRS... | 0.3911 | 0.2% | No | LRSK7_ORYSJ [Possible Remote
Homolog?]
9 | sp|Q8BMQ2|TF3C4_M... | 0.3950 | 0.1% | No | TF3C4_MOUSE [Possible
Remote Homolog?]
10 | sp|Q8K406|LGI3_MOUSE | 0.3971 | 0.9% | No | LGI3_MOUSE [Possible
Remote Homolog?]

```

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

```

```

-----
BLAST(Ref)   | 0.667         | 1   | 1.00         (Reference)

```

```

neural       | 0.4692        | 2   | 0.000

```

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

Method: neural

Rank | Neighbor ID | L2Dist | Ident. | In BLAST? | Bio Comment

1	sp Q09499 IRE1_CAEEL	0.3161	0.0%	No	IRE1_CAEEL [Possible Remote Homolog?]
2	sp Q5BCU8 ATG1_EMENI	0.3172	0.7%	No	ATG1_EMENI [Possible Remote Homolog?]
3	sp Q7XIT1 IRE1_ORYSJ	0.3234	1.1%	No	IRE1_ORYSJ [Possible Remote Homolog?]
4	sp Q4ICG8 LCMT1_G...	0.3247	0.7%	No	LCMT1_GIBZE [Possible Remote Homolog?]
5	sp P36002 PTK1_YEAST	0.3275	1.0%	No	PTK1_YEAST [Possible Remote Homolog?]
6	sp Q9GN97 PGPLD_D...	0.3284	0.0%	No	PGPLD_DROME [Possible Remote Homolog?]
7	sp Q68BL8 OLM2B_H...	0.3292	0.0%	No	OLM2B_HUMAN [Possible Remote Homolog?]
8	sp Q60M68 NDC1_CAEBR	0.3315	0.0%	No	NDC1_CAEBR [Possible Remote Homolog?]
9	sp E9QSG0 DSCA_ASPFU	0.3325	0.0%	No	DSCA_ASPFU [Possible Remote Homolog?]
10	sp P09706 US30_HCMVA	0.3366	0.0%	No	US30_HCMVA [Possible Remote Homolog?]