

RECOMMENDATION SYSTEM FOR FORUM



TS. Huỳnh Ngọc Tín Thầy Huỳnh Văn Tín

- Sinh viên thực hiện
 - 1. Dương Thị Hồng Hạnh 18520711
 - 2. Trần Thị Mỹ Linh 18520999





1. Task

User ID Recommender system Recommended post





https://vfo.vn/forum/











Diễn đàn Mới nhất Công nghệ Điện thoại	Máy tính Xe	Thủ Thuật	Hỏi đáp
Tin tức Máy Tính - Laptop	Chủ để	Bài viết	AMD xác nhận các trình điều k
	2.3K	10.1K	21:50 Hôm qua · Tin Nhanh
Phần cứng - Tư vấn mua Máy tính, Laptop	Chủ để	Bài viết	GTX 1060 Bị từ chối bảo hành
	5.6K	43.5K	10:56, Thứ năm · Tuấn Hà
Hòi đáp nhanh	Chủ đề	Bài viết	Xin review về Window 11
	21.8K	136.1K	14:28 · Trang Éc
Tin Học Căn Bản	Chủ để	Bài viết	Sửa lỗi SSL khi không truy cập
Chuyên mục con ▼	2.9K	8.8K	1/11/21 · thien1905
Kiến Thức Tin Học	Chủ đề	Bài viết	Hướng dẫn cập nhật Driver Ca
Chuyên mục con ▼	2.1K	23.5K	29/9/21 · Nguyenphuonganh290
Thủ thuật Windows	Chủ để	Bài viết	Topic thảo luận về hệ điều hà
	502	2.3K	16/10/21 · 0406
Download Phần Mềm Chuyên mục con ▼	Chủ để	Bài viết	Download phần mềm Hard Di
	3.2K	49.8K	20:38, Thứ tư · tt9
MAC OSX	Chủ đề	Bài viết	[Video] Cùng xem bản concep
	56	154	1/8/21 · tandobkhcm
Linux	Chủ đề	Bài viết	Đã có Ubuntu 20.10 chính thứ
	51	286	17/4/21 · 0406

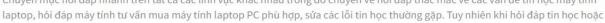




Hỏi đáp nhanh

Máy Tính >

Chuyên muc hỏi đáp nhanh trên tất cả các lĩnh vực khác nhau trong đó chuyên về hỏi đáp thắc mắc về các vấn để tin học máy tính



... 1092 Tiếp •

hỏi đáp về các lĩnh vực không được liên quan tới vi pham bản quyển hoặc các vấn đề nhạy cảm khác

Đã đọc Theo dỗi

Bô loc ▼

Viết bài

Tiêu để bài viết

Mời ae khoe góc làm việc máy tính

Trả lời: 742 Lươt xem:

> 30 16K

Trả lời:

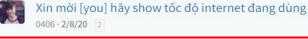
Lươt xem:

1/11/21 thien1905

9/10/21

minmin9

9/10/21



0406 · 2/8/20 2

0406 - 7/4/21

[Góc hỏi] Các thím bao lâu rồi vẫn chưa cài lai hê điều hành?

0406 · 4/3/21 2

Trả lời: 30 Lươt xem: 2K

minmin9





trên điện thoại khi rảnh?

[Hỏi nha các ban] Các ban hay chơi game gì nhất

Trả lời: Lươt xem:

86 93K

2/6/21 quyha1205

14:28



Trang Éc · 14:28

Thu mua itune giftcard Apple

Trả lời: Lươt xem: Trả lời:

0

Trang Éc



itunegiftcard · 22:27 Hôm qua

Lượt xem:

0 51

22:27 Hôm qua itunegiftcard

Bảo vê con thời 4.0 theo phương pháp 4.0 Thuphan001 · 21:44 Hôm qua

Trả lời: Lươt xem: 21:44 Hôm qua Thuphan001



POST.csv

Máy Tính > Hỏi đáp nhanh >

Nguyên nhân máy tính dùng ssd khởi động mất 20s?

minhvuong31011991 · 22/10/17

Theo dõi 🔲 #1

#1

quanity

 $Em\ d\`{u}ng\ m\'{a}y\ b\`{a}n, chip\ i3-4160, main\ H81-DS2, SSD\ Adata\ sp550\ 240GB\ (tr\~{o}ng\ 50\%), , ram\ 8GB., d\`{u}ng\ win\ 7\ 64\ bit.$

.

Lúc mới mua ổ SSD thì máy em khởi động chỉ mất 12s, (từ lúc ấn phím nguồn đến lúc đèn Num Lock sáng là 4s, còn 8s còn lại là load Win) & tắt máy chỉ 2s. Nhưng thời gian gắn đây em thấy máy khởi động tới 20s (mất 7s đèn Num Lock mới sáng, 13s load Win), tắt máy 6-7s.

٠

Em đã thử cài lại win, Ghost lại bản 1 năm trước, vệ sinh máy, tắt các chương trình khởi động cùng window, BenMark lại ổ SSD thì tốc độ vẫn như lúc mới mua, dọn file rác các kiểu, kiểm tra trong Bios (mà 1 năm nay e ko động vào Bios nên ko có chuyên bị chuyển qua IDE gì gì đó) nhưng vẫn không khắc phục được.

.

Ngày trước e cũng dùng SSD 120GB, sau 1 năm thì máy khởi động chậm hằn nên e mua ổ khác. Lúc đầu thì nhanh nhưng 1 năm sau thì vẫn bị tình trang ấy.

.

Bác nào có cao kiến gì chỉ em với. Em xin cảm ơn!

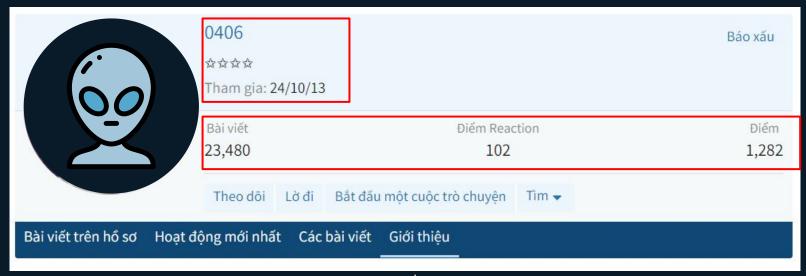
Sự phụ của ADMIN Mà 20s với 12s thì có sao đầu mà bạn so đo tính toán mãi, 8s đấy có đủ thời gian làm việc gì khác nữa đâu. Khi vào win, bật các phần mềm lên trong nháy mắt là OK rồi Báo xấu Like & Trá lời **3** #3 mình mua cái SSD cùi hồi đó có 1tr, 3-4 năm rồi dùng vẫn tốt, tốc độ giảm xuống kha khá nhưng ko snowolf901 sao. Ő SSD nó giảm tốc độ theo thời gian dùng mà. Tuyết Lang chênh lệch vài giây khởi động có đáng phải lo sốt vó lên thế ko ban. trải nghiệm dùng của bạn có bị ảnh hưởng gì nặng nề ko nè 🐽 Like & Trà lời Báo xấu Windows 7 thì hơi châm 1 tí, với lai ban dùng lâu rồi vây là phải rồi, thử vào BIOS khôi phục lai mặc latruor gthinh định xem có nhanh lên được không, nếu được cài windows 10 đi ban , bỏ màn hình boot khởi đồng Gà nhấ Forum khoảng 1~3s là lên màn hình rồi nhưng phải tắt cái update đi nhé **COMMENT.csv**

SSD bị già và chậm sau thời gian dài sử dụng, win cũ, các phần mềm sinh nhiều rác cũng là 1 nguyên

#2









1. Dataset



POST.csv
18.643 posts
10 features

COMMENT.csv
114.226 comments
5 features

USER.csv
10.997 users
8 features

not_available_user.txt
2391 userlinks

Removing inactive users :

POST.csv

COMMENT.csv







post=pd.read_csv('/content/drive/MyDrive/DS300/RS_DA/DATA_vfo/DATASET/Raw_data/POST.csv')
post

	id_post	title_post	url_post	views	comments	content_post	tags	id_author_post	url_author_post	datePublished_post
0	0	Mời ae khoe góc làm việc máy tính	https://vfo.vn//r/moi-ae-khoe-goc-lam-viec- may	663	4	\nMời ae khoe góc làm việc máy tính của mình t	0	179483	/m/0406.179483/	2021-04- 07T16:25:22+0700
1	1	Xin mời [you] hấy show tốc độ internet đang dùng	https://vfo.vn//r/xin-moi-you-hay-show-toc-do	16000	30	\nGói VNPT của mình đang dùng) còn các bạn mạ	['internet', 'mang', 'đo tốc độ']	179483	/m/0406.179483/	2020-08- 02T21:16:59+0700
2	2	[Góc hỏi] Các thím bao lâu rồi vẫn chưa cài lạ	https://vfo.vn//r/goc-hoi-cac-thim-bao-lau-roi	2000	30	\nMình thì hơn 3 năm nay rồi mà vẫn chưa 1 lần	['cài windows', 'hệ điều hành', 'hỏi đáp win']	179483	/m/0406.179483/	2021-03- 04T14:48:45+0700
3	3	[Hỏi nha các bạn] Các bạn hay chơi game gì nhấ	https://vfo.vn//r/hoi-nha-cac-ban-cac-ban-hay	93000	86	\nkhi rảnh thì mình hay thường chơi game Pubg	['giải trí', 'hỏi game', 'điện thoại']	179483	/m/0406.179483/	2019-01- 01T21:25:10+0700
4	11	Thiết kế khu vực nhà bếp	https://vfo.vn//r/thiet-ke-khu-vuc-nha-bep.143	98	2	\nChào mọi người, em đang tính decor lại khu b	['nha bep', 'thiết kế']	468274	/m/clarita.468274/	2021-10- 29T16:39:43+0700
	322	522	to:	22.00	1222	5.0	122	552	1922	522
18638	21813	Ai biết chia mạng thì giúp với	https://vfo.vn//r/ai-biet-chia-mang-thi-giup-v	6000	10	\ntính miền địa chỉ IP của subnet 2300 của lớp	['chế', 'của', 'lớp', 'với']	28594	/m/khung-troi- xanh.28594/	2011-05- 27T21:20:57+0700
18639	21814	Vấn đề giảm học phí	https://vfo.vn//r/van-de-giam-hoc-phi.8765/	4000	1	\nsinh viên đang học tại trường Đh công lập ở	['học', 'sinh viên']	17523	/m/loan_bl92.17523/	2011-05- 25T16:31:09+0700
18640	21815	Help! Driver am thanh	https://vfo.vn//r/help-driver-am-thanh.1412/	8000	7	\nmáy của em :\nmain boar : ZX- 865GVLM \ncó cà	['bạc', 'bạc liêu', 'cách', 'cần', 'chi', 'cho	233	/m/yono_lee.233/	2010-11- 06T12:39:43+0700
18641	21816	Hỏi về mạng ADSL	https://vfo.vn//r/ho%CC%89i-ve%CC%80- ma%CC%A3n	7000	11	\nmình ở cách xa bưu điện khoảng 4km mìn	['cách', 'cho', 'của', 'dow', 'download', 'gam	7384	/m/callofduty- 1982.7384/	2011-04- 19T18:10:32+0700
18642	21817	Cách bỏ dấu trong bài viết	https://vfo.vn//r/cach-bo-dau-trong-bai-viet.6	8000	5	\ncác pro cho hỏi. sao lúc trước mình viết bài	['bay', 'cách', 'cho', 'kho', 'microsoft', 'mi	7384	/m/callofduty- 1982.7384/	2011-04- 02T20:22:31+0700
186/13 ro	ws x 10 colu	imns								

18643 rows x 10 columns





cm=pd.read_csv('/content/drive/MyDrive/DS300/RS_DA/DATA_vfo/DATASET/Raw_data/COMMENT.csv')
cm

	id_post	id_author_cmt	url_author_cmt	content_cmt	datePublished_cmt
0	0	179483	/m/0406.179483/	lem:lem:lem:lem:lem:lem:lem:lem:lem:lem:	2021-06-25T12:03:23+0700
1	0	179483	/m/0406.179483/	\n\nguongtp:\n\n\n\t\t\tBan dùng máy cấu hìn	2021-06-25T12:05:06+0700
2	0	466251	/m/clone-vui-tinh.466251/	Ai cũng Win10, có mình tui Win7	2021-06-25T17:42:27+0700
3	0	468305	/m/thien1905.468305/	\n\n\n\nmình có cái bàn làm việc thật rộng	2021-11-01T09:10:10+0700
4	1	392333	/m/lammori.392333/	\n\n0406:\n\n\n\t\t\tVào app My viettel xem	2020-09-25T10:56:51+0700
	8410	3	100	740	620)
114221	21817	4	/m/t.4/	Đề: Cách bỏ dấu trong bài viết\n\nbạn chọn bản	2011-04-03T12:48:02+0700
114222	21817	7384	/m/callofduty-1982.7384/	Đề: Cách bỏ dấu trong bài viết\n\nđược rồi mìn	2011-04-03T19:57:39+0700
114223	21817	2	/m/le-minh.2/	Đề: Cách bỏ dấu trong bài viết\n\nVì mình đã b	2011-04-02T21:05:29+0700
114224	21817	7384	/m/callofduty-1982.7384/	Đề: Cách bỏ dấu trong bài viết\n\nvẫn không đư	2011-04-03T12:21:07+0700
114225	21817	39	/m/yehyeh.39/	Đề: Cách bỏ dấu trong bài viết\n\nsao bỏ cái p	2011-04-10T08:43:00+0700

114226 rows x 5 columns

user=pd.read_csv('/content/drive/MyDrive/DS300/RS_DA/DATA_vfo/DATASET/Raw_data/USER.csv')
user

	id_user	name_user	star	DateJoin	DateLastAccess	comment_count	Reaction_Score	Score
0	179483	0406	1.0	2013-10-24T16:03:00+0700	NaN	23480	99	1132
1	468274	Clarita	1.0	2021-10-29T16:13:29+0700	2021-11-02T09:03:36+0700	1	0	1
2	460618	thien-ha-bet	1.0	2020-07-13T16:17:58+0700	2021-11-01T16:53:54+0700	7	0	1
3	468272	laptoptitan.vn	1.0	2021-10-29T10:29:04+0700	2021-10-29T10:48:30+0700	1	0	1
4	468270	Phúc Lucas	1.0	2021-10-29T09:42:46+0700	2021-10-31T15:20:36+0700	2	0	1
	***			555%	345*	***	***	
10992	51414	hoangpupu	NaN	2011-07-11T01:40:44+0700	2012-11-21T01:55:12+0700	37	0	0
10993	33680	::Rùa::	NaN	2011-05-26T19:21:00+0700	2013-02-27T23:20:36+0700	337	0	0
10994	20	^^KuTi_K_	NaN	2010-08-11T11:30:00+0700	2014-08-24T15:57:53+0700	379	0	0
10995	348	NhocShjnky	NaN	2010-11-20T22:42:19+0700	2010-11-20T22:54:18+0700	1	0	0
10996	39	yehyeh	NaN	2010-08-19T19:32:00+0700	2011-04-19T22:42:45+0700	14	0	0

10997 rows x 8 columns

2. Approach



Content based

Neural + Collaborative

O1 Text
Preprocessing

02 Embedding

O3 Clustering

LDA - topic modeling

5 Similarity calculation



1. Text preprocessing

- Text data preparation:
 - Lowercase letter
 - Unicode Normalization
 - Diacritics Normalization

Remove Stopword

Tokenization

Clean text:

- Teencode
- Punctuation
- Image, URL,...

2. Embedding

• TF-IDF:

TfidfVectorizer(max_features=15000)
fit on : title, content, tags.

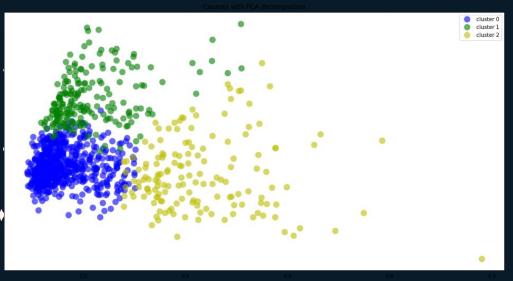
• W2V:

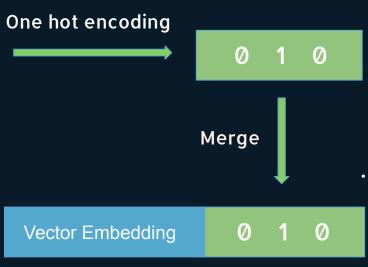
vector_size=200, window=2, epoch=50
vector_size=300, window=2, epoch=30

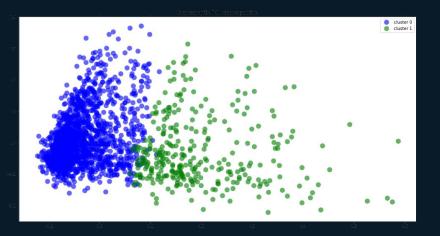
PhoBERT:

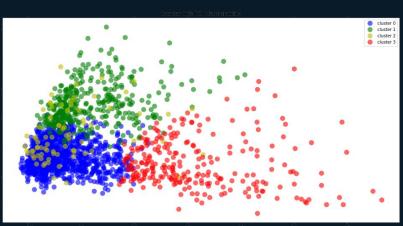
vinai/phobert-base

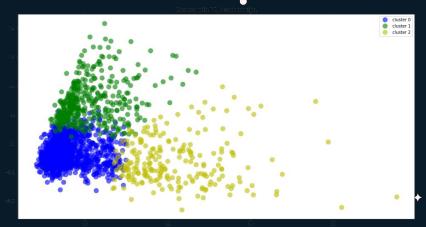


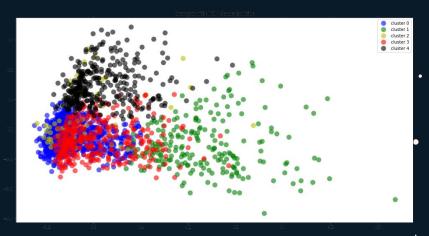














Latent Dirichlet Allocation

Topics

gene	0.04
dna	0.02
genetic	0.01

life	0.02
evolve	0.01
organism	0.01

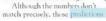
brain	0.04
neuron	0.02
nerve	0.01

data	0.02
number	0.02
computer	0.01
111	

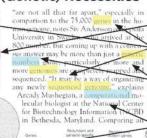
Documents

Topic proportions and assignments











SCIENCE • VOL. 272 • 24 MAY 1996

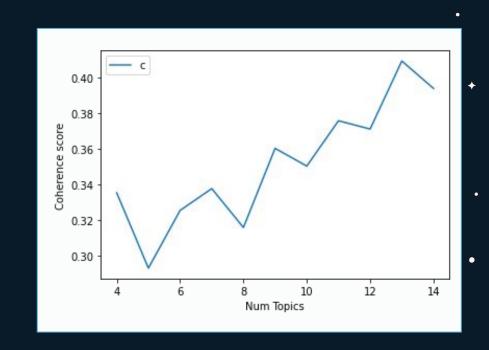


4. LDA

Coherence Value

passes=40, terations=200, chunksize = 10000

Title + tags apply TF-IDF





```
'0.058*"phần mềm" + 0.050*"video" + 0.036*"key" + 0.033*"youtube" + 0.031*"gỡ" + 0.030*"idm" + 0.022*"help" + 0.022*"me" + 0.018*"nhac" + 0.016*"hoat đông"'),
'0.029*"google" + 0.028*"chrome" + 0.028*"chrome" + 0.028*"chrome" + 0.024*"giúp đỡ" + 0.024*"trình duyệt" + 0.024*"font" + 0.022*"link" + 0.021*"tải" + 0.019*"lỗi" + 0.018*"photoshop"),
'0.027*"gta" + 0.017*"tin hoc" + 0.016*"v" + 0.016*"tai sao" + 0.015*"chip" + 0.014*"exe" + 0.012*"thông tin" + 0.012*"lỗi" + 0.011*"flash" + 0.011*"hê thống"),
'0.029*"office" + 0.029*"of cứng" + 0.025*"ram" + 0.024*"lỗi" + 0.022*"nâng cấp" + 0.022*"o" + 0.021*"thắc mắc" + 0.021*"file" + 0.017*"dữ liệu" + 0.017*"xóa"'),
'0.035*"xanh" + 0.031*"cài đát" + 0.027*"ứng dung" + 0.022*"phát" + 0.022*"iphone" + 0.022*"màn hình" + 0.020*"lỗi" + 0.019*"main" + 0.019*"pc" + 0.019*"chǎn"'),
'0.068*"tu văn" + 0.053*"mua" + 0.047*"laptop" + 0.024*"asus" + 0.022*"dell" + 0.021*"bàn phím" + 0.018*"active" + 0.016*"pass" + 0.015*"mấy" + 0.015*"máy tính"'),
'0.064*"khởi động" + 0.040*"virus" + 0.027*"đen" + 0.024*"máy tính" + 0.023*"diệt" + 0.023*"màn hình" + 0.018*"phần mềm" + 0.018*"chương trình" + 0.017*"triệu" + 0.016*"hỏi đáp"'),
'0.071*"usb" + 0.038*"boot" + 0.024*"đĩa" + 0.021*"lỗi" + 0.020*"win" + 0.019*"pin" + 0.018*"sửa" + 0.017*"laptop" + 0.017*"cài" + 0.016*"ổ đĩa"').
'0.031*"cpu" + 0.022*"net" + 0.020*"window" + 0.019*"liên minh" + 0.019*"pascal" + 0.018*"tốc đô" + 0.016*"kiểm tra" + 0.014*"anh em" + 0.014*"viết" + 0.014*"cũ"'),
'0.043*"máy tính" + 0.040*"tắt" + 0.029*"tự động" + 0.026*"máy" + 0.022*"bản quyền" + 0.021*"bật" + 0.019*"online" + 0.016*"lỗi" + 0.015*"laptop" + 0.014*"biểu tương"),
'0.065*"game" + 0.059*"màn hình" + 0.050*"card" + 0.035*"cấu hình" + 0.019*"trơ giúp" + 0.018*"vga" + 0.018*"máy tính" + 0.017*"icon" + 0.017*"chay" + 0.016*"lỗi"'),
'0.060*"wifi" + 0.049*"giúp" + 0.041*"mang" + 0.028*"kết nối" + 0.023*"lỗi" + 0.020*"đổi" + 0.019*"bắt" + 0.018*"viêt" + 0.018*"mât khẩu" + 0.016*"trang"'),
'0.088*"win" + 0.063*"cài" + 0.041*"lỗi" + 0.035*"driver" + 0.034*"ghost" + 0.029*"xp" + 0.026*"facebook" + 0.024*"giúp" + 0.024*"windows" + 0.022*"update" )]
```



5. Similarity calculation

Profile

ping liên minh huyền thoại đôi 300ms khó_chịu ...

hiện_tại you trình_duyệt suốt thời_kỳ sử_dụng .

2 tư_vấn laptop tầm triệu tiêu_đề mua láp tầm tr...

3 hộ thông_số vga bình_thường chẳng tự_nhiên gam...

4 giúp_đỡ modem wifi adsl tenda d152 vnpt kết_nố...

Title + Content + Tags

Title + Tags

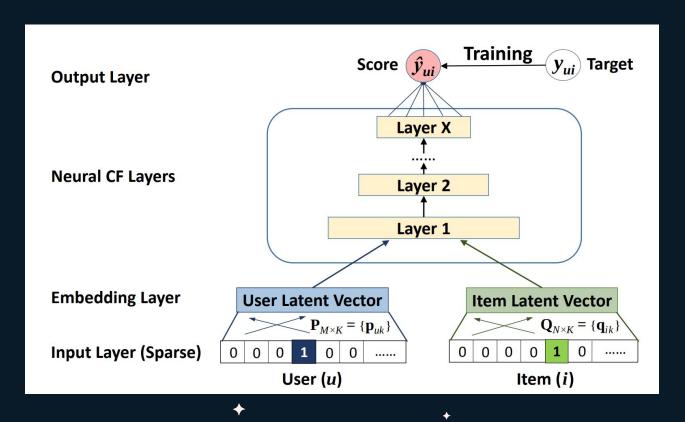
Cosine

P

Testset

	Unnamed: 0	id_user	id_post	label					
50399	50399	386961	5510	1					
50400	50400	386961	19648	0					
50401	50401	386961	19961	0					
50402	50402	386961	11255	0					
50403	50403	386961	183 <u>5</u> 3	0					
		***	***	***					
50495	50495	386961	20632	0					
50496	50496	386961	17499	0					
50497	50497	386961	20530	0					
50498	50498	386961	1921	0					
50499	50499	386961	10765	0					
101 row	101 rows × 4 columns								

Neural + Collaborative



3. Evaluation





4. Results

RS						NDCG
			Aggregate Text	Aggregate Profile	THE TALLO	NDCG
			Title + tags	post	0.319	0.262
	TF-IDF			cmt	0.215	0.165
				post + cmt	0.225	0.174
СВ			Full	post	0.319	0.277
				cmt	0.243	0.206
				post + cmt	0.261	0.212
		cluster		cmt	0.336	0.271
		LDA	Title + tags	post + cmt	0.162	0.110
	W2V	D200 + cluster	Full	cmt	0.334	0.280
		D300 + cluster	run		0.312	0.269
	PhoBert		Full	cmt	0.216	0.152
\mathbf{CF}	NCF	G	MF	amt	0.266	0.250
CF	NOI	Ne	euMF	m cmt	0.250	0.210

Table 1: Bảng kết quả thực nghiệm thu được.



* THANKS FOR WATCHING

