

Изучение профилей русскоязычных сообществ LiveJournal: региональный аспект

© А.В. Сычев

Воронежский государственный университет
sav@cs.vsu.ru

Аннотация

В докладе представлены результаты исследования на основе данных из профилей сообществ блог-хостинга *LiveJournal*. Рассматриваются такие задачи как реконструкция хронологии создания сообществ, вычисление корреляции между атрибутами профилей. Наибольшее внимание в работе было уделено интересам сообществ. Был изучен характер распределения интересов в профилях, как путем расчета простейших числовых характеристик, так и с помощью процедуры кластеризации. Для сравнения приведены результаты исследования регионального распределения читателей в профилях сообществ.

1 Введение

Социальные сети в WWW, привлекающие сегодня к себе всеобщее внимание пользователей Интернета, сформировались за очень короткий промежуток времени (буквально 5 - 8 лет). Они объединяют в себе блоги (сетевые дневники), сети медиа-ресурсов, сети персональной информации (MySpace, LinkedIn, Facebook, «В контакте.ru», «Одноклассники.ru» и другие), системы закладок (del.icio.us), wiki-энциклопедии и другие. Количество пользователей в этих сетях увеличивается с беспрецедентной скоростью, вызывая вполне обоснованный интерес к ним у представителей науки, бизнеса и ИТ-индустрии.

Социальные сети фактически представляют собой гигантское хранилище общедоступной информации, в первую очередь, персонального характера. Анализ этой информации может дать ценные сведения о структуре современного информационного общества и процессах, протекающих в нем.

За рубежом данная область является предметом

интенсивных исследований, что подтверждается многочисленными публикациями по этой тематике. Направление «Социальные сети» было выделено отдельно в рамках такой авторитетной международной конференции как *World Wide Web* в 2008 и 2009 годах.

Хороший обзор актуального состояния исследований, подходов, инструментов и приложений, связанных с блогосферой, приведен в [1]. В этой работе рассмотрены следующие направления исследований: моделирование блогосферы, кластеризация, сбор данных, влияние и распространение информации, доверие и репутация, фильтрация спам-блогов.

Результаты исследований на примере конкретных коллекций данных приведены в работах [2-5]. В [5] представлены результаты широкомасштабного исследования на основе данных из 4 популярных социальных сетей *Flickr*, *YouTube*, *LiveJournal* и *Orkut*. Общий объем данных составил 11.3 млн. пользователей и 328 млн. связей между ними. Как показало исследование, социальные сети структурно отличаются от ранее исследованных сетей, в частности Веб. Для них характерна существенно большая доля симметричных связей и гораздо более высокий уровень локальной кластеризации. В [3] была исследована и проведена экспериментальная оценка (на основе данных из базы DBLP) предложенной авторами модели кластеризации, использующей помимо атрибутов объектов информацию о связях между ними. В [4] предложена методика FacetNet, основанная на байесовской оценке максимального правдоподобия, комбинирующая задачи выявления сообществ и их эволюции в рамках единого процесса. Для экспериментальной оценки была использована база DBLP. Феномен микроблоггинга на примере социальной сети *Twitter* рассмотрен в [2]. В этой работе приведены результаты исследования топологических и географических характеристик этой сети.

Русскоязычная блогосфера исследована в очень малой степени, при этом публикации, содержащие серьезные структурные исследования этой социальной сети, практически отсутствуют. Отдельные аспекты русскоязычной блогосферы

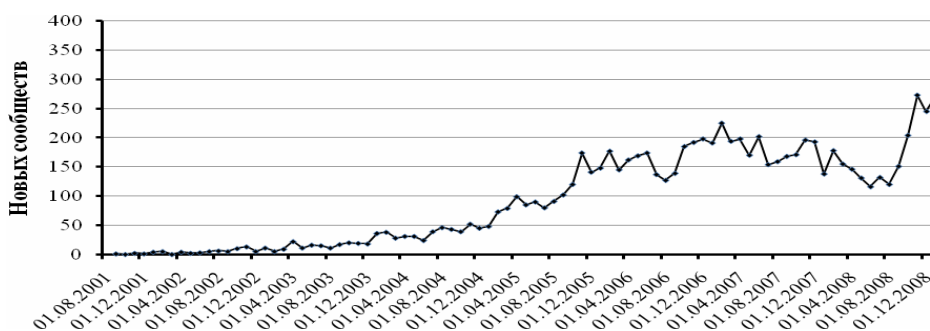


Рисунок 1. Хронология создания сообществ.

Таблица 1. Усредненные значения атрибутов профилей русскоязычных сообществ *LiveJournal*.

Атрибут	1	2	3	4	5	10	11	12	13	14
Значение	22	1,5	0,4	131	126	219	10	69	1,0	1763

Таблица 2. Корреляция между атрибутами профилей сообществ.

	2	3	4	5	6	7	8	9	10	11	12	13	14
1	0,09	0,12	0,12	0,12	-0,06	-0,07	0,24	0,04	0,06	0,21	0,23	0,21	0,04
2		0,21	0,17	0,17	-0,11	-0,12	0,12	0,04	0,15	0,12	0,17	0,12	0,15
3			0,14	0,16	0,01	0,02	0,09	0,05	0,04	0,11	0,10	0,11	0,04
4				0,95	-0,13	-0,29	0,13	0,10	0,72	0,18	0,13	0,09	0,63
5					-0,13	-0,27	0,13	0,09	0,59	0,20	0,15	0,10	0,51
6						0,61	0,14	-0,03	-0,09	-0,01	-0,21	0,05	-0,06
7							0,22	-0,06	-0,21	-0,02	-0,29	0,03	-0,17
8								0,03	0,08	0,14	0,26	0,19	0,05
9									0,13	0,02	0,02	0,04	0,15
10										0,15	0,05	0,06	0,90
11											0,17	0,14	0,14
12												0,16	0,02
13													0,05

регулярно освещаются в отчетах компании Яндекс (раздел портала “Исследования Яндекса”, <http://company.yandex.ru/researches/>), например в [6].

В работе [7] уже были представлены результаты расчета характеристик для сообществ блог-хостингов *LiveJournal* и *LiveInternet*, полученных на основе атрибутов профилей сообществ. Там же приведены графики и таблицы, полученные в результате кластеризации сообществ и их интересов.

В данной работе приведены новые результаты исследования профилей русскоязычных сообществ блог-хостинга *LiveJournal* (www.livejournal.com). Во-первых, было существенно увеличено количество сообществ - вместо прежних примерно трех тысяч русскоязычных сообществ *LiveJournal* - теперь для исследования были использованы 8870 профилей сообществ. Во-вторых, появился новый атрибут – *регион*. Так в 1898 профилях в качестве

региона была указана Российская федерация, в 575 – бывший СССР, а в остальных – регионы РФ.

Ввиду ограничений поискового сервиса *LiveJournal*, удалось получить неполный список сообществ по РФ и Москве (только первые 2000). По всем остальным регионам списки сообществ были получены полностью. Данное ограничение, безусловно, вносит определенные коррективы при анализе распределений атрибутов профилей и их характеристик, но, как представляется, несущественно влияет на результаты анализа связей между всеми остальными регионами РФ.

Целями данного исследования являются определение того, насколько информативным может быть анализ только данных из профилей сообществ, исследование региональной специфики русскоязычных сообществ *LiveJournal* и изучение возможности использования данных из профилей сообществ для анализа связей между регионами.

Таблица 3. Рейтинг регионов РФ по числу сообществ (представлены первые 10 регионов)

Регион	Сообществ	Среднее значение атрибута профиля сообщества													
		1	2	3	4	5	6	7	8	10	11	12	13	14	
0	1898	31	1,8	0,6	345	331	0,5	18.06.2007	18.02.2009	593	24	119	1,4	5397	
1	1733	20	1,5	0,3	79	76	0,4	27.09.2006	20.06.2008	108	6	67	0,9	442	
2	1146	30	1,7	0,6	164	162	0,6	10.08.2007	15.01.2009	289	17	92	1,3	2321	
3	575	26	1,6	0,3	139	133	0,2	06.04.2006	07.03.2008	159	7	73	0,8	1582	
4	294	15	1,4	0,2	38	38	0,3	22.08.2006	15.01.2008	67	3	34	0,7	429	
5	241	15	1,4	0,3	24	23	0,6	17.04.2007	16.05.2008	52	3	38	0,8	258	
6	195	16	1,4	0,2	31	30	0,4	17.10.2006	09.03.2008	54	4	34	0,9	201	
7	190	13	1,5	0,2	28	29	0,5	05.12.2006	11.04.2008	36	3	36	0,7	120	
8	186	13	1,5	0,3	39	32	0,5	09.11.2006	15.01.2008	82	4	32	0,6	650	
9	156	15	1,4	0,2	42	43	0,4	28.09.2006	21.04.2008	72	3	45	0,9	240	
10	145	14	1,4	0,3	30	30	0,5	13.01.2007	30.04.2008	43	4	43	0,8	106	

Таблица 4. Корреляция между количеством сообществ в регионе и средним значением атрибутов профиля сообщества. В первой строке учтены все регионы, включая “РФ”, во второй – исключен регион “РФ”, в третьей – исключены “РФ”, Москва, Санкт-Петербург и бывший СССР.

атрибут	1	2	3	4	5	6	7	8	10	11	12	13	14
корреляция	0,42	0,44	0,20	0,82	0,80	-0,11	-0,05	0,34	0,77	0,70	0,48	0,25	0,76
	0,32	0,34	0,11	0,67	0,63	-0,16	-0,12	0,19	0,55	0,49	0,34	0,17	0,53
	0,08	0,34	-0,07	0,32	0,25	-0,27	-0,26	-0,07	0,31	0,23	0,10	0,02	0,28

2 Исследование атрибутов профилей русскоязычных сообществ LiveJournal

2.1 Хронология создания сообществ

На графике, представленном на рисунке 1, отражена хронология формирования русскоязычных сообществ в блог-хостинге LiveJournal. На нем четко просматриваются сезонные колебания скорости появления новых сообществ с характерными весенними пиками и летними спадами

2.2 Атрибуты профилей сообществ

Также было проведено изучение характеристик атрибутов профилей сообществ. В таблице 1 приведены усредненные значения для атрибутов профилей, а в таблице 2 - корреляция между атрибутами профилей сообществ. Номерами в них обозначены атрибуты: 1 - количество интересов, 2 - количество смотрителей, 3 - количество модераторов, 4 - количество участников, 5 - количество читателей, 6 - тип аккаунта, 7 - дата создания, 8 - дата последнего обновления, 9 - количество вирт. подарков, 10 - количество записей, 11 - количество меток, 12 - количество избранных записей, 13 - количество картинок пользователя, 14 - количество полученных комментариев.

Анализ корреляции между атрибутами профилей сообществ (таблица 2) показал тесную связь между количеством участников (4) и количеством читателей (5), между количеством записей (10) и количеством комментариев на них (14).

Прослеживается также связь между количеством участников (читателей) и количеством записей (10) и комментариев (14) в сообществах.

2.3 Региональное распределение сообществ и характеристик их атрибутов

В таблице 3 приведены количество сообществ и средние значения атрибутов профилей для первой десятки в рейтинге регионов (по числу сообществ): 0 - РФ, 1 - Санкт-Петербург, 2 - Москва, 3 - бывший СССР, 4 - Новосибирская область, 5 - Московская область, 6 - Нижегородская область, 7 - Республика Татарстан, 8 - Самарская область, 9 - Свердловская область, 10 - Ростовская область.

Значения корреляции между количеством сообществ в регионе и средними значениями атрибутов профилей этих сообществ представлены в таблице 4. Обращает на себя внимание заметная зависимость средних значений числа читателей, участников, записей и комментариев от числа сообществ в регионе.

3. Кластеризация регионов по интересам

Как представляется, наиболее информативным является анализ распределения интересов и

пользователей (читателей), указанных в профилях сообществ, по регионам.

В исследовании не ставилась задача поиска наиболее эффективных для данной задачи алгоритмов кластеризации регионов по сообществам, скорее преследовалась цель получения группировок регионов в первом приближении, поэтому выбор конкретного алгоритма был достаточно произвольным.

В работе был использован метод аггломеративной кластеризации Ланса-Уильямса. Первичное расстояние между сообществами рассчитывалось по формуле:

$$\rho(c_1, c_2) = \frac{|c_1 \cap c_2|}{\sqrt{|c_1|} \cdot \sqrt{|c_2|}}$$

Регион c_i рассматривался как множество интересов, указанных в профилях сообществ этого региона. При проведении процедуры кластеризации расстояние между кластерами рассчитывалось по формуле среднего расстояния.

В качестве исходных данных для процедуры кластеризации регионов была использована матрица "регион-интерес", из которой была сформирована матрица "регион-регион".

На основе матрицы "регион-регион" были получены кластеры регионов при различных значениях порога кластеризации $Th = 1/2^k$, $k = 0, 1, 2, 3, 4, 5, 6$.

Таблица 5. Невырожденные кластеры регионов (по интересам) при значении порога $Th = 0.125$

Кластер	Регион	Сообществ	Интересов
6	Астраханская область	20	297
	Псковская область	17	124
7	Башкортостан Республика	91	1066
	Свердловская область	156	1929
13	Вологодская область	26	408
	Новосибирская область	295	3794
	Самарская область	186	2010
35	Москва	1165	22571
	Санкт-Петербург	1737	22596
38	Нижегородская область	195	2530
	Ростовская область	145	1790
44	Пермский край	103	1104
	Чувашская Республика	16	173
45	Приморский край	88	1205
	Хабаровский край	49	783

Таблица 6. Невырожденные кластеры регионов (по интересам) при значении порога $Th = 0.0625$.

Кластер	Регион	Сообществ	Интересов
3	Алтайский край	22	242
	Бурятия Республика	6	29
5	Архангельская область	22	196
	Астраханская область	20	297
	Башкортостан Республика	91	1066
	Белгородская область	25	221
	Владимирская область	17	305
	Волгоградская область	53	483
	Вологодская область	26	408
	Воронежская область	96	1015
	Ивановская область	12	77
	Иркутская область	105	894
	Калининградская область	110	1679
	Карелия Республика	18	387
	Кемеровская область	40	386
	Костромская область	9	115
	Краснодарский край	112	1483
	Курская область	9	44
	Липецкая область	17	154
	Москва	1165	22571
	Московская область	241	3034
	Нижегородская область	195	2530
	Новосибирская область	295	3794
	Омская область	86	678
	Приморский край	88	1205
	Псковская область	17	124
	Ростовская область	145	1790
	Рязанская область	42	637
	Самарская область	186	2010
Санкт-Петербург	1737	22596	
Саратовская область	66	814	
Сахалинская область	16	238	
Свердловская область	156	1929	
Ставропольский край	122	1072	
Татарстан Республика	190	2109	
Томская область	43	781	
Тюменская область	61	784	
Удмуртская Республика	34	455	
Хабаровский край	49	783	
Челябинская область	122	1443	
13	Курганская область	7	79
	Чукотский автономный округ	1	5
16	Марий Эл Республика	8	164
	Новгородская область	29	351
22	Пермский край	103	1104
	Чувашская Республика	16	173

Таблица 7. Список интересов, указанных в профилях сообществ из всех регионов, входящих в кластер №45 при значении порога $Th=0.125$.

№	Интерес	№	Интерес	№	Интерес	№	Интерес
1	35 мм	38	дизайн	74	мужчины	111	репортаж
2	art	39	дожди	75	музыка	112	ресторан
3	arthouse	40	дружба	76	музыканты	113	Россия
4	canon	41	женщины	77	набережная	114	самореализация
5	depth of field	42	живопись	78	наркотики	115	секс
6	foto	43	жизнь	79	Находка	116	семья
7	linux	44	загадки	80	негатив	117	слайд
8	minolta	45	здоровье	81	недвижимость	118	сон
9	nikon	46	зенит	82	новости	119	спорт
10	olympus	47	зеркалка	83	ночные клубы	120	старый город
11	pentax	48	игра	84	общение	121	стихи
12	photo	49	иллюстрации	85	огонь	122	счастье
13	photography	50	интересные люди	86	Оле Нидал	123	съемка
14	pinhole	51	интернет	87	оптика	124	творчество
15	slr	52	информация	88	отдых	125	театр
16	unix	53	искусство	89	панк	126	тигры
17	автомобили	54	история	90	пиво	127	трамваи
18	алкоголь	55	Карма Кагью	91	пленка	128	троллейбусы
19	архитектура	56	Кармапа	92	политика	129	учеба
20	барабаны	57	катастрофа	93	помощь	130	философия
21	бесплатно	58	кафе	94	портрет	131	флирт
22	брусчатка	59	кино	95	портфолио	132	фонтаны
23	буддизм	60	Китай	96	Поэзия	133	фото
24	Буддизм	61	классика	97	поэзия	134	фотоаппарат
25	бытовая техника	62	книги	98	Приморье	135	фотовыставки
26	вино	63	книги по фотографии	99	Природа	136	фотографии
27	Владивосток	64	компьютеры	100	природа	137	фотография
28	водка	65	кофе	101	провинция	138	Фотография
29	встречи	66	красота	102	программирование	139	фотожурналистика
30	выдержка	67	лето без воды	103	прогулки	140	фотопечать
31	гламур	68	литература	104	проза	141	художники
32	город	69	любовь	105	психология	142	цвет
33	горячая вода	70	люди	106	путешествия	143	цифровая фотография
34	графика	71	макро	107	работа	144	черно-белая фотография
35	гулять	72	медитация	108	радость	145	шашлыки
36	деньги	73	море	109	развлечение	146	Япония
37	диафрагма			110	реинкарнация		

В таблицах 5 и 6 приведены невырожденные кластеры (т.е. содержащие более одного элемента) регионов, в которых расстояние между регионами (отражающее общность интересов, указанных в профилях сообществ) оказалось не ниже порога Th .

Первые невырожденные кластеры образуются при значении $Th = 0.125$. Далее по мере снижения порога происходит резкое расширение единственного суперкластера при уменьшении общего количества невырожденных кластеров, что и отражено в таблице 8.

Таблица 8. Характеристики невырожденных кластеров регионов (по интересам).

Порог Th	0.125	0.0625	0.03125	0.015625
Кол-во кластеров	7	5	4	3
Средний размер кластера	2,2	9,2	15	16,5
Медиана размера кластера	2	2	2,5	3

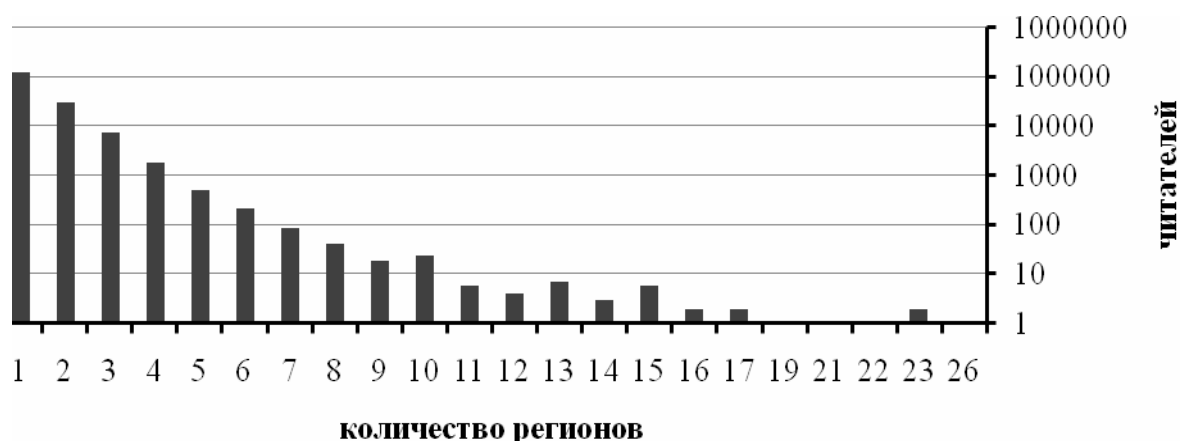


Рисунок 2. Распределение читателей по количеству регионов

Таблица 9. Характеристики невырожденных кластеров регионов (с выборкой интересов из средней части рангового распределения).

Порог Th	0.125	0.0625	0.03125	0.015625
Количество кластеров	4	9	15	10
Средний размер кластера	2,0	4,0	7,2	13,3
Медиана размера кластера	2	2	2	3

Список общих интересов для одного из кластеров, содержащего два региона (Приморский край и Хабаровский край), приведен в таблице 7.

Анализ частотного распределения интересов в профилях региональных сообществ показал, что в большинстве регионов доминирующим по частоте употребления в профилях сообществ является название областного (краевого, республиканского) центра, далее по рангу может встречаться название региона. Высокую частоту также имеют общераспространенные интересы, например: музыка, фото, компьютеры и т.п. В целом ранговое распределение интересов по частоте напоминает распределение Ципфа, что и объясняет характер кластеризации регионов по интересам. Однако согласно принципу Луна (Luhn), наиболее специфичными в распределении Ципфа являются элементы, находящиеся в средней части рангового распределения. Как представляется, исключение из анализа самых высокочастотных и самых редких (т.е. неспецифических) интересов может дать более содержательные результаты при кластеризации регионов по интересам. В таблице 9 приведены характеристики кластеров регионов, полученных для матрицы “регион-регион”, построенной для выборки интересов, имеющих частоту в интервале от 2 до 20.

4. Кластеризация регионов по читателям

Для исследования были использованы списки читателей сообществ, указанных в их профилях. В

19-ти самых крупных сообществах представленное в списке количество читателей отличалось от реального (до полутора раз).

Общее количество читателей (уникальных) составило порядка 160 тысяч.

На рисунке 2 представлено распределение читателей по количеству регионов, журналы сообществ из которых они читают. Например, всего лишь один пользователь является читателем блогов сообществ из 26 регионов, а число читателей блогов сообществ из не более чем одного региона составило порядка 100 тысяч.

В таблице 11 дается более детальное представление о распределении читателей по регионам. В ней указано, сколько пользователей являются читателями журналов сообществ только одного региона, двух и т.д. Результаты кластеризации регионов по читателям существенно отличаются от кластеризации по интересам. В таблице 10 представлены характеристики полученных кластеров (невырожденных).

Таблица 10. Характеристики невырожденных кластеров регионов (по читателям).

Порог Th	0,125	0,0625	0,03125	0,015625
Кол-во кластеров	1	2	7	16
Средний размер кластера	2	2,5	2,4	2,7
Медиана размера кластера	2	2,5	2	2

При значении порога $Th = 0.125$ был выделен единственный невырожденный кластер, содержащий Москву и Санкт-Петербург. Кластеры для других значений порога Th приведены в таблице 12.

Заключение

В данной работе была рассмотрена возможность использования ограниченного объема данных

Таблица 11. Распределение читателей по регионам.

	Всего	Только 1 регион	Только 2 региона				Только 3 региона	Только 4 региона	Только 5 регионов
			Регион+ Москва	Регион+ Спб	Регион+ другой	Всего			
Читателей	219969	122618	25104	20762	16380	62246	21939	7424	2560
%	100	56	11	9	7	28	10	3	1

Таблица 12. Невырожденные кластеры регионов (по читателям) при различных значениях порога Th .

$Th = 0.0625$

Кл-р	Регион	Читателей
20	Калининградская область	3760
	Москва	96520
	Санкт-Петербург	60616
41	Томская область	559
	Новосибирская область	5952

$Th = 0.015625$

Кл-р	Регион	Читателей
1	Адыгея Республика	62
	Башкортостан Республика	1143
3	Алтайский край	373
	Ростовская область	2597
5	Тюменская область	1212
	Архангельская область	612
	Вологодская область	517
7	Мурманская область	790
	Белгородская область	481
8	Владимирская область	487
	Брянская область	462
9	Пермский край	1456
	Курганская область	106
11	Бурятия Республика	94
	Воронежская область	1463
15	Нижегородская область	3504
	Томская область	559
	Челябинская область	2348
	Новосибирская область	5952
17	Иркутская область	1915
	Новгородская область	444
	Калининградская область	3760
	Москва	96520
	Санкт-Петербург	60616
	Свердловская область	4023
	Краснодарский край	2088
25	Ленинградская область	601
	Удмуртская Республика	569
29	Костромская область	318
	Липецкая область	236
	Мордовия Республика	9
30	Тамбовская область	62
	Чувашская Республика	256
34	Татарстан Республика	3131
	Московская область	3634
35	Пензенская область	166
	Саратовская область	724
38	Приморский край	3422
	Хабаровский край	646
40	Самарская область	2941
	Тверская область	378
44	Сахалинская область	463
	Ханты-Мансийский автономный округ	130

$Th = 0.03125$

Кл-р	Регион	Читателей
1	Адыгея Республика	62
	Башкортостан Республика	1143
7	Белгородская область	481
	Владимирская область	487
18	Калининградская область	3760
	Москва	96520
	Санкт-Петербург	60616
	Свердловская область	4023
33	Краснодарский край	2088
	Мордовия Республика	9
38	Тамбовская область	62
	Новосибирская область	5952
42	Томская область	559
	Пензенская область	166
44	Саратовская область	724
	Приморский край	3422
	Хабаровский край	646

(только информации, извлекаемой из профилей сообществ пользователей) для изучения региональной структуры русскоязычной блогосферы. В рамках данной задачи наиболее информативными атрибутами из профилей сообществ являются списки интересов и читателей (участников).

Для анализа связей между регионами была использована процедура кластеризация.

Конечно, приведенные результаты выглядят скорее как набор статистических фактов, а использованные в работе общие подходы и их реализации в конкретных алгоритмах требуют более детального анализа и дальнейшей проработки. Тем не менее, исходя из этих результатов, можно сделать ряд интересных выводов.

Так, результаты кластеризации регионов по интересам и по читателям проявляют разные аспекты связи между регионами. Негеографические (экономические, культурные и подобные им) связи могут выявляться в большей степени с помощью кластеризации по интересам.

Кластеризация по читателям отражает в большей степени географическую близость регионов (включая и степень мобильности пользователей между регионами), причем, в большей степени это заметно для регионов, находящихся в неевропейской части РФ. Повышение уровня экономического и технологического развития региона приводит к конвергенции этих двух аспектов межрегиональных связей.

Можно провести аналогию с методами, которые применяются в поисковых системах для гипертекстовых документов. Для них используются 2 взаимно дополняющих друг друга подхода: содержательный (документ – как множество терминов) и топологический (документ – как узел гипертекстового графа). В случае регионов (а также отдельных сообществ и даже пользователей) эквивалентом термина является интерес, а топологическая связь между ними реализуется через отдельных читателей (участников) сообщества (или друзей пользователя). Как представляется, использование данной аналогии позволило бы применять хорошо проработанные методы анализа и поиска документов для поиска суперсообществ (регионов), сообществ и отдельных пользователей в социальных сетях.

Представленные в работе подходы могут быть использованы как для изучения связей между регионами, так и для построения поискового образа конкретного региона.

Литература

- [1] N. Agarwal, H. Liu. "Blogosphere: Research Issues, Tools, and Applications"//SIGKDD Explorations, 2008, Vol. 10, Issue 1. - pp. 18 - 31.
- [2] A. Java, X. Song, T. Finin, B. Tseng "Why we twitter: understanding microblogging usage and

communities" // WebKDD/SNA-KDD '07: Proceedings of the 9th WebKDD (2007), pp. 56-65.

- [3] R. Ge, M. Ester, B. J. Gao, Z. Hu, B. Bhattacharya, B. Ben-Moshe "Joint Cluster Analysis of Attribute Data and Relationship Data: The Connected k-Center Problem, Algorithms and Applications". ACM Transactions on KDD, Vol. 2, No. 2. (2008), Article 7.
- [4] Y.-R. Lin, Y. Chi, S. Zhu, H. Sundaram, B. L. Tseng "Analyzing Communities and Their Evolutions in Dynamic Social Networks" // ACM Trans. Knowl. Discov. Data, Vol. 3, No. 2. (2009), pp. 1-31.
- [5] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, B. Bhattacharjee "Measurement and Analysis of Online Social Networks"// Proceedings of the 7th ACM SIGCOMM conference on Internet measurement, 2007, San Diego. pp. 29-42.
- [6] Состояние блогосферы российского интернета. По данным поиска по блогам Яндекса. Весна 2008 г. [Электрон. ресурс] – Режим доступа: (http://download.yandex.ru/company/yandex_on_blogosphere_spring_2008.pdf)
- [7] Сычев А.В., Гадебский И.А. Изучение характеристик сообществ русскоязычной блогосферы //Электронные библиотеки: перспективные методы и технологии, электронные коллекции: Труды Десятой Всероссийской научной конференции «RCDL'2008». – Дубна: ОИЯИ, 2008. – С. 200-2009.

The Study of The Profiles Features for Russian Blog Communities Hosted at LiveJournal: Regional Aspect

A.V. Sychev

On the base of data driven from profiles of russian communities, hosted at LiveJournal, some research tasks were carried out, like: new communities emerging chronology reconstructing, profiles attributes correlation estimating and Russian Federation regions clustering using both interests list and readers list from communities profiles.