

зировать с помощью статей словаря анализа. Если в результате анализа какой либо из словоформ получены грамматические характеристики, не совпадающие с характеристиками, которые диктует для данной словоформы статья словаря синтеза, то считается, что конвертирование было проведено неправильно либо статья словаря синтеза неправильная, и данная словоформа выделяется красным цветом. Пользователь может подтвердить правильность статей словаря анализа и тем самым добавить их в дерево-образ формируемого словаря анализа. Для удобства использования программы при проверке большого числа статей пользователь может активировать опцию автоматического подтверждения. При условии, что автоматическая проверка не выявила ошибок, эта опция приводит к автоматическому подтверждению правильности всех статей всякий раз, когда пользователь щелкает по кнопке «Дальше». Если активировать опцию «Пропускать проверенное», то при каждом щелчке по кнопке «Дальше» будет выбрана следующая не проверенная статья словаря синтеза.

Результатом работы является удобное и надежное средство получения словарей, используемых для морфологического анализа и синтеза. С помощью этих словарей можно эффективно решать проблемы автоматического перевода, извлечения информации из текста, автоматического индексирования баз данных в информационно поисковых системах, сжатия текстовых данных, проверки грамматической правильности текста, создания электронных словарей и обучающих систем, и на дальнейших этапах исследований.

Литература

1. Вейник П. А., Волосевич А. А. Структура грамматического словаря для синтеза русских словоформ, *Инженерный вестник*, 2006, 1(21)/3, с.174 - 177.
2. Вейник П. Проблема расстановки ударений в русском тексте, *Материалы конференции «Информационные системы и технологии»*, Часть 2, Мн., 2006.
3. Зализняк А. А. *Грамматический словарь русского языка*, М., 1983.
4. Кулагина О.С. *Морфологический анализ русских именных словоформ*, М., 1986.
5. Кулагина О. С. *Морфологический анализ русских глаголов*, М., 1986.

РОЛЕВЫЕ МОДЕЛИ В СОЦИАЛЬНЫХ СЕТЯХ

Зяц Ю.Э.
Беларусь, Минск

Введение

Социальная сеть представляет собой множество социальных объектов, связанных между собой некоторой формой социальных отношений. В связи с высокой популярностью таких сервисов, как MySpace [1], LiveJournal [2], YouTube [3], Flickr [4] социальные сети стали представлять интерес для исследования. В настоящее время данные сервисы широко используются для блоггинга, подкаст-вещания, публикации данных, поиска сотрудников и работода-

телей и т.д., аудитория крупнейших из них составляет более 10^7 пользователей [5]. Блогосфера активно используется для PR, партизанского маркетинга и продвижения товаров.

Модель

Большинство исследователей онлайн-социальных сетей используют либо общедоступные данные, собранные поисковыми пауками на социальных сайтах [6], либо различные модели, основанные на сгенерированных случайным образом значениях [7]. Для получения достоверных и полных данных (таких как личная переписка, косвенная пользовательские данные, хронология изменения данных и т.д.), которые не могут быть представлены в публичных источниках по причинам приватности, была создана система для развертывания социальных сетей, включающая в себя шесть доменов второго уровня (www.gamefabrique.com, www.sureloads.com и другие). В настоящий день нагрузка сети составляет 2500 уникальных пользователей в день (хитов), 22500 просмотров (хостов), общее количество зарегистрированных пользователей – более 2000, полученных сообщений, публичных и приватных – 400 и 1000 соответственно. Программное обеспечение позволяет пользователям вести блог, публиковать фотографии, создавать личные страницы, комментировать (древовидная система комментирования, позволяющая сопоставлять ответы пользователей на конкретное сообщение) и обмениваться сообщениями через систему внутренней почты (PM). Кроме этого, пользователи могут отмечать других пользователей как «друзей» (система френдования, аналогичная LiveJournal). Общие сведения о пользователе, статистика активности, списки друзей и взаимных друзей отображаются в профиле пользователя.

Каждый пользователь социальной сети характеризовался по двум параметрам – вовлеченность (социальная адаптация) (s) и самореклама (p) на основе статистики его профиля, комментариев и личных сообщений.

	Вовлеченность	Самореклама
Повышающие факторы	<ul style="list-style-type: none"> • Количество полученных сообщений в PM • Количество отправленных комментариев с полученными ответами • Количество созданных страниц • Наличие аватара • Наличие фотографий • Количество взаимных друзей 	<ul style="list-style-type: none"> • Количество отправленных комментариев без полученных ответов • Количество отправленных сообщений в PM без полученных ответов • Количество друзей • Количество элементов в списке желаний выше порогового

Низкая вовлеченность и низкий уровень саморекламы. Недавно зарегистрировавшиеся пользователи, либо пользователи, чье отношение к социальной сети невозможно определить.

Высокая вовлеченность и низкий уровень саморекламы. К этой группе относятся большинство пользователей социальной сети. Количество взаимных друзей у таких пользователей возрастает постепенно с течением времени и со-

относится с активностью в социальной сети (рис. 1, пользователи JGCD, jakemawkes).

Низкая вовлеченность и высокий уровень саморекламы. В эту группу с высокой вероятностью попадают экспериментаторы-новички, использующие для поднятия собственного рейтинга и получения социального распознавания (social recognition) различные инструменты социализации (рис. 1, пользователь treutado). Кроме того, сюда были отнесены фрилансеры, помогающие формировать контент социальной сети. Практически все пользователи отмечали у фрилансеров малое количество факторов вовлеченности, что, в свою очередь, было причиной их нулевого френдования.

Высокая вовлеченность и высокий уровень саморекламы. Свойственны лидерам сообществ и в некоторой степени – модераторам и владельцам узлов, на которых развернута социальная сеть. При этом высокий уровень вовлеченности достигается благодаря занимаемой социальной роли, а уровень саморекламы – благодаря количеству опубликованного контента.

В упрощенном виде формула для расчета выглядит следующим образом:

$$S = \sum_{i=1}^n \frac{(k_i * param_i - threshold_i)}{param_{average}}$$

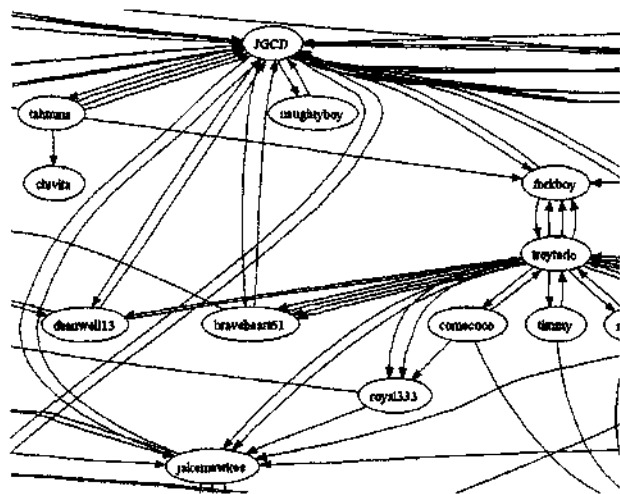
где k_i , $threshold_i$ – коэффициент и пороговое значение фактора соответственно

Аналогичная формула применяется для расчета значения p .

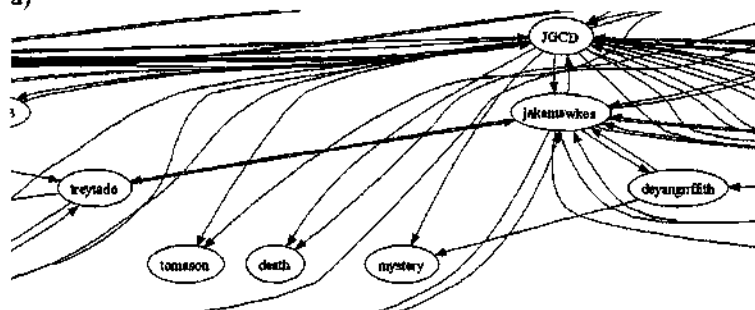
Процентное распределение пользовательских профилей по выделенным моделям показано в таблице 1.

Таблица 1. Распределение пользовательских профилей по уровням вовлеченности и саморекламы.

Вовлеченность (s)	Самореклама (p)				
	0..0,25	0,25..0,5	0,5..0,75	0,75..1	
0..0,25	0,914	0,005	0,008	0,012	
0,25..0,5	0,001	0,003	0,008	0,008	
0,5..0,75	0,005	0,008	0,001	0,006	
0,75..1	0,014	0,005	0,002	0,003	



a)



b)

Рис. 1. Френдование (a) и PM (b) пользователей JGCD ($s=1$, $p=1$), jakemawkes ($s=1$, $p=0.25$) и treyado ($s=0.25$, $p=1$), фрагмент

Выводы

Как видно из статистики, приведенной при описании модели, значительная часть коммуникаций в социальной сети приходится на private отношения (а в социальных сетях, построенных на основе реальных социальных систем, таких как МойКруг [8] – private оффлайновые отношения), что часто затрудняет, а то и делает невозможным установление полной картины развития отношений между пользователями.

При создании социальных сервисов необходимо давать пользователям возможность преодоления начальной сегрегации. Так, например, введение системы отображения аватаров и дней рождения новых пользователей, увеличило не только процент заполняемости профилей, но и число пользовательских регистраций, несмотря на усложнение регистрационной формы (добавление обяза-

тельного поля в регистрационную форму для некоторых аудиторий способно уменьшить число регистраций на 20-30%[9]) (рис. 2).

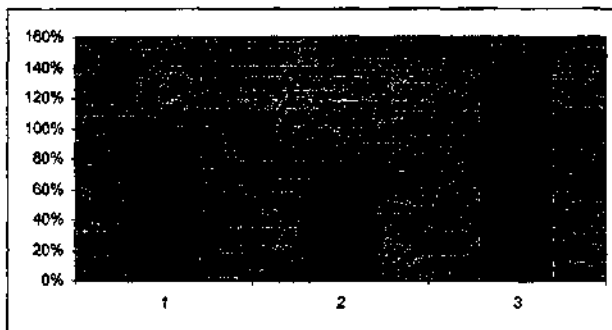


Рис. 2. Нормализованное отношение количества регистраций при введении дополнительных социальных сервисов. 1-первоначальное значение, 2-при усложнении регистрационной формы, 3-после добавления сервисов

Использование значений p и z целесообразно при фильтрации пользователей, рассылающих рекламные сообщения и спам.

Литература

1. MySpace [Электрон.ресурс] / MySpace, Inc. – Beverly Hills, CA, 2007 – Режим доступа: <http://www.myspace.com>. – Дата доступа 01.09.2007
2. LiveJournal.com [Электрон.ресурс] / LiveJournal, Inc. – San Francisco, 2007 – Режим доступа: <http://www.livejournal.com>. – Дата доступа 10.04.2007
3. YouTube - Broadcast Yourself [Электрон.ресурс] / Google, Inc. – Herndon, VA, 2007 – Режим доступа: <http://www.youtube.com>. – Дата доступа 01.09.2007
4. Welcome to Flickr - Photo Sharing [Электрон.ресурс] / Yahoo!, Inc. – Sunnyvale, CA, 2007 – Режим доступа: <http://www.flickr.com>. – Дата доступа 01.09.2007
5. Wikipedia – MySpace [Электрон.ресурс] / Wikimedia Foundation, Inc. – Saint Petersburg, Florida, 2007 – Режим доступа: <http://en.wikipedia.org/wiki/MySpace>. – Дата доступа 10.04.2007
6. Is Britney Spears Spam? [Электрон.ресурс] / М. Е. J. Aaron Zinman – MIT Media Lab, 2007 – Режим доступа: <http://www.smg.media.mit.edu/papers/Zinman/britneyspears.pdf>. – Дата доступа 01.09.2007
7. Random graph models of social networks [Электрон.ресурс] / М. Е. J. Newman, D. J. Watts, S. H. Strogatz – Columbia University, 2002 – Режим доступа: http://www.pnas.org/cgi/content/full/99/suppl_1/2566. – Дата доступа 01.09.2007
8. Online-секция на Конференции разработчиков компьютерных игр 2007. [Электрон.ресурс] / ООО «КРИ» – Москва, 2007 – Режим доступа: http://kriconf.ru/2007/rec/KRI_2007_BusinessDevelopment_06apr_Saturn_OnLine_p1.ogg. – Дата доступа 01.09.2007