# **Chapter 9 Applied Methods and Techniques for Modeling and Control on Micro-Blog Data Crawler**

Kai Gao, Er-Liang Zhou and Steven Grover

Abstract Models can provide mechanisms to improve system performance. This chapter presents the applied methods and techniques for modeling and controlling on micro-blog crawler. With the rapid development of social studies and social network, millions of people present or comment or share their opinions on the platform everyday, and as a result, produce or spread their opinions and sentiments on different topics. The microblog has been an effective platform to know or mine social opinions. In order to do so, crawling the relevant microblog data is necessary. But it is hard for a traditional web crawler to crawl micro-blog data as usual, as by using Web 2.0 techniques such as AJAX, the micro-blog data is dynamically generated rapidly. As most microblogs' official platforms cannot offer some suitable tools or RPC interface to collect the big data effectively and efficiently, we present an algorithm on modeling and controlling on micro-blog data crawler based on simulating browsers' behaviors. This needs to analyze the simulated browsers' behaviors in order to obtain the requesting URLs to simulate and parse and analyze the sending URL requests according to the order of data sequence. The experimental results and the analysis show the feasibility of the approach. Further works are also presented at the end.

Keywords Models · Social networks · Micro-blog · Crawler

S. Grover

K. Gao (🖾) · E.-L. Zhou

School of Information Science and Engineering, Hebei University of Science and Technology, No. 26 YuXiang Road, Shijiazhuang, 050000, Hebei, China e-mail: gaokai@hebust.edu.cn

Comrise Company, Concord Center Building 2, Hazlet, NJ, 07730, USA e-mail: steven@comrise.com

#### 9.1 Background

Nowadays, most of the research in the fields of mechatronic systems or social studies have spent significant effort to find rules from various complicated phenomena by principles, observations, measured data, and logic derivations. The rules are normally summarized as concise and quantitative expressions or "models," and this can provide mechanisms to improve the system (represented by its model) performance. As for the social studies, the social network data (e.g., Twitter, Facebook, Sina\_microblog, Tecent\_microblog, etc.) have attracted millions of users and academic and industry researchers to research on modeling and mining the knowledge behind the magnanimity information, and as a result, there has been tremendous interest in social networks. Due to its fast development and wide usage, the microblog has attracted the attention of users, enterprises, governments, and researchers, and so applied methods and techniques for modeling and controlling in this field is very important.

As the foundation of the micro-blog data mining, data collection is the key phase, crawling or collecting the relevant micro-blog data effectively and efficiently is important. But the microblog has many differences compared with traditional web applications. For example, there are many online users, at the same time, its different interactive and displaying mode and login operation are needed, and AJAX technology is widely used, etc. Traditional web crawlers, for instance, can only get the corresponding web pages, but they cannot get the relevant structure and the corresponding social relationships as well as users' backgrounds and fans. That is to say, being different from traditional web application, there are some differences on micro-blog data's login operation, display way, privacy policy, data processing, etc. So the traditional web crawler is not suitable for micro-blog data crawling or collection.

This section presents some details on modeling of micro-blog data crawler based on simulating browsers' behaviors. On the basis of this method, we have collected several million blog data in a short time period.

### 9.2 Motivation

Although there has been some research on AJAX-based web pages, the technique is not suitable to the micro-blog application. Encouraging developers to develop applications on micro-blog services, some providers of micro-blog services usually offer some special APIs, which can provide developers with the probability of constructing uniform and universal architecture to utilize the APIs to automatically download and save these special data. But the mere APIs-based method has some limitations on rights, calling times, special policies, and so on, and some extra tasks cannot be done by only using these official APIs.

In this chapter, we present some strategies based on simulating browsers' behaviors to obtain the data from micro-blog platform. The main idea is to simulate browsers'

behaviors by using the browser's (e.g., FireFox) core to get the corresponding data. This can solve the problem of parsing the JavaScript code, and can do special login operation, etc. In order to crawl the data effectively, we present the following strategies: (1) focused crawling on some special crowds; (2) meta-topic searching and crawling: that is to say, we crawl the special contents by using the microblog's searching function; (3) parallel crawling: based on big data processing by using the *Redia* and *MongoDB*, we use the multiprocessing technology to download and save the data simultaneously. The proposed crawler is composed of four modules, i.e., simulating module, data crawling module, data parsing module, and data persistence module. The experimental results and the analysis show the feasibility of the approach. Further works are also presented at the end.

#### 9.3 Related Work

Online social networking technologies enable individuals to simultaneously share information with any number of peers. With the launch of Twitter in 2007, the microblog has become highly popular, and many researchers want to investigate the micro-blog information propagation patterns [1] or analyze structures of the microblog network to identify influential users [2]. Reference [3] discusses some of the ways in which earlier works used text content to analyze online networks, as well as background on language coordination and the exchange-theoretic notions of power from status and dependence. Reference [4] studies several long-standing questions in media communications research, in the context of the micro-blog service Twitter, regarding the production, flow, and consumption of information. A framework which enriches the semantics of Twitter Messages (i.e., tweets) and identifies topics and entities (e.g., persons, events, products) mentioned in tweets is present in reference [5]. Reference [6] conducts a study on recommending URLs posted in Twitter messages and compares strategies for selecting and ranking URLs by exploiting the social network of a user as well as the general popularity of the URLs in Twitter. Authors of reference [7] investigate the attributes and relative influence of 1.6M Twitter users by tracking 74 million diffusion events that took place on the Twitter follower graph over a 2-month interval, and they conclude that the word-of-mouth diffusion can only be harnessed reliably by targeting large numbers of potential influencers, thereby capturing average effects. Reference [8] examines the role of social networks in online information diffusion with a large-scale field experiment, and the authors further examine the relative role of strong and weak ties in information propagation. Although stronger ties are individually more influential, it is the more abundant weak ties that are responsible for the propagation of novel information, and the authors suggest that weak ties may play a more dominant role in the dissemination of information online than currently believed. In reference [9], authors address the problem of discovering topically meaningful communities from a social network, and authors propose a probabilistic scheme that incorporates topics, social relationships, and nature of posts for more effective community discovery, and then

they demonstrate the effectiveness of the model and show that it performs better than existing community discovery models. Reference [10] examines the application of an event-driven sampling approach to the Live Journal social network, and the approach makes use of the "always on" atom feed provided by Live Journal that contains all public blog posts in near real-time to inform the sampling process of user friendship networks, and this has the effect of targeting sampling toward the public active users of the network. In addition to proposing models and algorithms for learning the model parameters and for testing the learned models to make predictions, reference [11] develops techniques for predicting the time by which a user may be expected to perform an action.

As for data crawling, in order to overcome the inherent bottlenecks with the traditional crawling, reference [12] proposes the design of a parallel migrating web crawler. Reference [13] proposes a dynamic data crawling methods, which include the sensitive checking of website changes and dynamic retrieving of pages from target websites, and the authors implement an application and compare the performance between the conventional static approaches and the proposed dynamic ones. In reference [14], authors present a novel URLs ordering system that relies on a cooperative approach between the crawlers and the web servers based on file system and web log information, and the proposed algorithm is based on file timestamps and web log internal and external counts. Reference [15] presents a micro-blog service crawler named as *MBCrawler*, which is designed on the APIs provided by micro-blog services, and the architecture is modular and scalable, so it can fit specific features of micro-blog services. Reference [16] presents a dynamic cooperation model for different crawlers' message exchanging, and both the experimental results and the application validate the feasibility of the algorithm.

As for the modeling methods, reference [17] presents the commonly used statistical modeling methods, such as stepwise regression, radial basis function partial least squares, partial robust M-regression, ridge regression, and principal component regression that can be applied in the proposed multicollinearity domain. The Viterbi algorithm, a widely used maximum likelihood estimating method, can be used in natural language processing, and reference [18] presents an effective search space reduction for human pose estimation with Viterbi algorithm.

Although the proposed algorithm has some relationship with the above related work, there are many differences. The proposed crawler works with simulating browser behavior to collect Sina\_Micro-blog (http://weibo.com/) and Tecent\_Micro-blog (http://t.qq.com/) data. The proposed algorithm is based on simulating browsers' behaviors. As for browser, reference [19] describes how to calculate various object-oriented metrics of three versions of Mozilla Firefox, and the neural network approach can predict high and medium severity errors more accurately than low severity errors.

The experimental results and the analysis show the feasibility of the proposed approach.

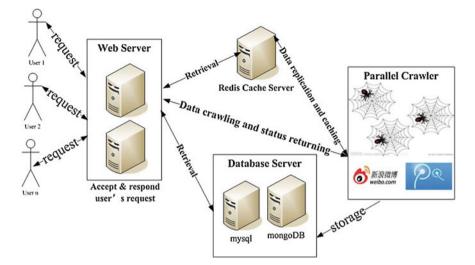


Fig. 9.1 System architecture

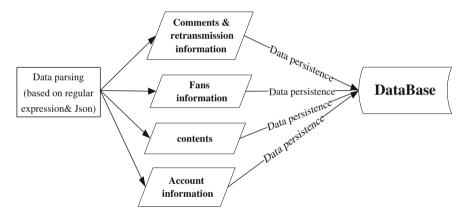


Fig. 9.2 Parsed data and its persistence

### 9.4 System Architecture

Social networks are often huge, and therefore crawling the micro-blog data could be both challenging and interesting. As microblog's big data properties, it is impossible to crawl all the micro-blog data. Instead, it is feasible to crawl some kinds of data (e.g., account information, contents or topics, attentions or fans, etc.). In this section, we propose the system architecture on parallel crawling. Figure 9.1 shows the architecture, and Fig. 9.2 shows the parsed data and its persistence.

In practice, we can use the *RDBMS* to store the parsed data, and the *Redis* is used as the cache server, so users' retrieval request can be done through the web server.

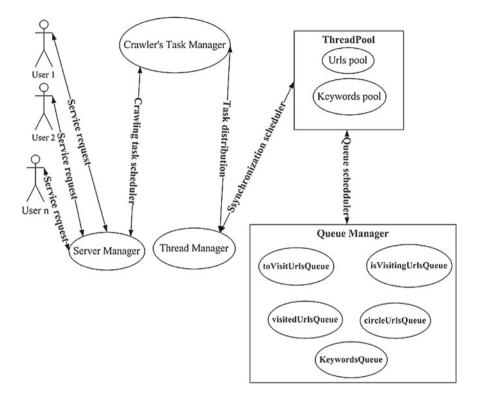


Fig. 9.3 Multi-thread-based parallel crawling

In detail, as for the multi-thread based parallel crawling, we use a thread pool and a queue manager to schedule the tasks. There are several different queues, including the *toVisitUrlsQueue*, *isVisitingUrlsQueue*, *visitedUrlsQueue*, *circleUrlsQueue*, *keywordsQueue*, etc., see Fig. 9.3.

# 9.5 Case Studies and Implementation of the Simulated Browser-Based Crawling

Instead of merely using the official APIs, we propose a simulated browser-based crawling, as the merely APIs-based method has some limitations on rights, calling times, and so on, and perhaps some extra tasks cannot be done by only using official APIs. We present some strategies based on simulating browsers' behaviors to obtain the micro-blog data, and the proposed crawler is composed of four modules, i.e., simulating module, data crawling module, data parsing module, and data persistence module.

(a) Started	Time Chart					Constant -	Time Chart				-	
		!	P	Time	Sen	Started				P	Time	Sen
	Tencent Webo_Voice from you, Echo from	the world			_		Tencent Webo_Voice from yo	ou, Echo from t	the world			
+ 0.000				0.139	114	+ 0.000					0.128	
+ 00:00:00.184	Home TecentWebo	0.139 -> 1		0.139	114		Home TecentWebo	1	0.120 -1	_	0.128	
	Home_recentivebo			E 170	818		Home_TecentWebo				5.282	
•							Home_TecentWebo	1		-	0,162	
- In	and the second s					4	Home_recentiveso	1		•	0.202	2.47
Overview Time	Chart Headers Cookies Cache	Query String	POST	Data	onteni	• 1						
Parameter	Value		_			Overview Time	Chart Headers Cookie	Cache	Query String	POST	Data	Conten
action	9-42-1380163779265					Parameter	Value	1				
aid	46000101		- 1			action	3-34-1380163936991	*				
daid	6		- 1			aid	46000101					
dumy	loginerroralert		- 1			daid	6					
fp from_ui	lognerroraiert		- 1			dumy						
mom_u			- 1			Þ	loginerroralert					
9			- 1			from_ui	1					
is_type			- 1				1					
is_ver	10047		- 1			1.00	1					
login_sig	IgVyuGiStB2RcOoUo3NUXdxhU7rTXcP	L-HzFmZSw7Pc1	7WHe	RETG"USXO	RCWV	s_type	1					
low_login_enable	1	59990 N 5985562	1	1		s_ver	10047					
low_login_hour	720		- 1			ogin_sig	IgVyuGiStB2RcOoUo3NL	KdxhU7rTXcPj	jL-HeFm2Sw7Po	17WHg8	QXzU*p1	ISRCWV
mbao_css			- 1			ow_login_enable	1					
p	41561389AD8DF36A92170237950A57	CA	- 1			ow_login_hour	720					
ptlang	1033		- 1			nbao_css	E4F595884495F8E5D39		-			
ptredirect	1		- 1			otlang	1033	291900030200	8°			
pt_rsa	0		- 1			otredirect	1					
regmaster			- 1			pt_rsa						
t	1		- 1			regmaster						
u	623143506@qq.com		- 1				1					
ul	http://t.qq.com		- 1				623143506@og.com					
verifycode	IXYG					1	http://t.gg.com					
						verifycode	IOUF					

Fig. 9.4 Different situations with the same account

## 9.5.1 Simulation of the Login Operation and Cookies Data Obtaining

Commercial websites often use technologies (e.g., HTTP compression, SSL encryption and chunked encoding) to provide some reasonable levels of security and system performance. As for the micro-blog data effectively crawling, simulated login operation is necessary. Otherwise (for example, by only using official APIsbased crawling), only few data can be crawling. Here, the simulated login operation means this kind of operation allows the crawler to use some legal accounts and their corresponding passwords to login the corresponding micro-blog platforms, and the key phase is the encrypted data parsing. Here, we use the HttpWatch [20], which integrates with Internet Explorer or Mozilla Firefox to provide some unrivaled levels of HTTP monitoring, without the need for separately configured proxies or network sniffers. Simply interacting with a website, HttpWatch can display a log of requests and responses alongside the web page itself, and it can even show interactions between the browser and its cache. As a result, each HTTP transaction can be examined or parsed so as to see the values of headers, cookies, query strings, and other HTTP-related data. HttpWatch can work well with these technologies to provide a view of HTTP activity. By using HttpWatch, we can obtain 21 or more different parameters during the simulated login phase. But in practice, there usually exist some different situations during the login phase, and Fig. 9.4 shows the two situations when using the same account.

(a)		(b)		
Started	Time Chart ! P	00:00:05.488 Tencent Welbo_Voice from you, Echo from the world	0.042	458
	Tencent Webo_Vaice from you, Echo from the world !	O0:00:09.380 Tencent Webo_Voice from you, Echo from the world		
= 00:00:00.175 + 0.000	Home_TecentWeb0	+ 0.000	0.039	817
+0.031		0.039 -> !	0.039	817
+ 0.330 + 0.578		00:00:12.695 Tencent Webo_Voice from you, Echo from the world	0.144	1164
+0.570	/ /	00:00:12.882 Home_TecentWebo	1.515	8825 2
1		00:00:14.508 Home TecentWebo	2.189	5447
Overview Time	Chart Headers Cookies Cache Query String PO:			
Parameter	John	4		
aid	46000101			
daid f_url	6	Overview Time Chart Headers Cookies Cache Query String POST Data	Content	Stream V
Later	0	application/x-javascript; charset=utf-8 : 60 bytes	1	
low_login_hour	720			
nodirect	0	ptui_checkVC('0' '!ODQ1.'\x00\x00\x00\x25\x24\x6a\x52';		
ptiang	1033			
ptredirect	301			
philip	YSWdxyK8RHq2187CzCcsLr8Tyb3ORDEfco3kTupW8C0_			
pttype	1 0.000 0.000			
regnaster	0			
service	login			
s_url	http://t.qq.com	verifyCode uin ninCode		
un	623143506	din pricode		

Fig. 9.5 Requested preliminary parameters (a) and the return values after the requested period (b)

		ript; charset=utf-8 :										Q Find the gopy of cost 3
	iew Time	Chart Headers		Cache: 4	Query String	POST-Data	Content	Stream	Warni	ngs (17) C	omment	
										-		
00	100:14.508	Home_Tecent//ebo				1	2.189	5447	5704	(21 requests	0	
00:	:00:12.882	Home_TecentWebo				1	1.515	8825	21724	(44 requests	0	
					0.1	66 -+ t	0.144	1164	1867	1 request		
	+ 0.000					1	0.144	1164	1867	CET 33	200	Http://bfogin2.og.com/login?u=623143506%40og.com8p=87839D8C72D12F02757886C8E367F4921
001	00:12.695	Tencent Webo Voic	e from you, i	icho from th	binow tel							

Fig. 9.6 Returned results

From the parsed results, we can conclude that the requested URLs usually contain static and dynamic parameters (e.g., the parameter *P* and *verifycode* in Fig. 9.4, and the *verifycode* parameter is usually used as the password encryption). As the microblogs' login passwords are usually multilevers and multilayer encrypted, it usually contains some other preliminary parameters, and the displayed parameter in Fig. 9.5b (i.e., "x00x00x00x25x24x6ax52") is the encrypted parameter in Fig. 9.4a. Now the encrypted resolving phase is finished, and the returned or parsed content is shown in Fig. 9.6.

As for judging whether the corresponding user is legal or illegal, it needs to analyze the cookies data. On the other hand, whether actually login or not, when requesting the server data, if the user can get the legal cookies, he or she can obtain the same data as if he or she really "login" the web server. In detail, in order to obtain the cookies data, it needs three steps. First, it needs to obtain the *verifycode* and *uin* parameters, see algorithm 1 below. Second, by using the JavaScript analysis engine and invoking the encrypted function, we can obtain the parsed parameters, see Fig. 9.7. Last, it needs to merge the relevant parameters to obtain the corresponding cookie data, which is the result of the simulating login phase, see algorithm 2 below, and the parsed cookies data result is shown in Fig. 9.8.

```
//parsing the verifyCode and uin parameters
Algorithm 1 (i.e., void getCheckVC())
Input:
(1) preLoginUrl//requested URLs before the login operation;
(2) username
(3) host//parameter on requesting the preLoginUrl
Output:
(1) retJson: returned parameters on preLoginUrl request
(2) verifyCode
(3) uin
Step1. retJson=getPreLoginJson(preLoginUrl,username,host);
Step2. verifyCode=parseVerifyCode(retJson);
Step3. uin=parseUin(retJson);
```

```
//obtaining the cookie data
Algorithm 2 (i.e., void getCookies())
Input:
(1) loginUrl // Url to login the corresponding micro-blog page
(2) username
(3) password
(4) verifyCode //verification code obtained from the former step
(5) host
(6) referrer// means the source urls to do the login operation
Output: cookie data
Step1. cookie=getCookies(loginUrl,username,password,verifyCode,
host,referer);
Step2.cookie=checkValidate(cookie) //verify the cookie
Step3.transmitCookie(cookie)//transfer the obtained cookies to
the corresponding threads
```

#### 9.5.2 Data Parsing and Persistence

After collecting the corresponding data, it needs to be parsed and stored. As for the content, there are some differences between data within the blogger's main page and other common pages. As usual, the main page returns data in a traditional way, while other common pages usually use AJAX [21] and JSON [22] technology to return data to client in order to enhance the performance or optimize the user experience. By using JavaScript, these returned data can be parsed and filled into the corresponding sites. Figure 9.9 shows the private crawled data, and algorithm 3 shows the main steps of this processing step.

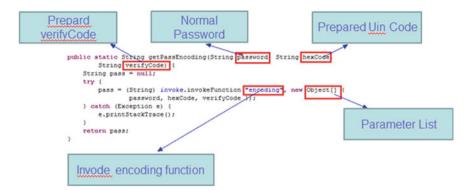


Fig. 9.7 The parsed parameters



Fig. 9.8 The cookies data of the simulating login phase

Started	Time Chart	1 P	Time	Sent	Received	Method	Result	Type	URL																		
00:00:00.000	EvaLongoria_Tencent Webo																										
+ 0.000		1	1.504	1242	21970	GET	200	. 8	http://t.qq.com/EvaLongoriavip																		
+ 1.173			0.001	0	0	GET	(Cache)	0	phtp://mat1.gbmg.com/www/mb/css/style_																		
+ 1.416	11		0.004	0	0	GET	(Cache)	-	http://mat1.atima.com/www/mb/ma/v1/b2																		
4						III	/																				
a l let				1		1																					
Overview Time	Chart Headers Cookies Cache	Query String POST Data	Content	Stream	Warm	ngs (2)	omment																				
ext/html; charset=u	utf-8:84619 bytes, gzip compressed to 2166	69 bytes ( 74.4 % saving )		-	S																						
	That amaginal of heads"here	//url on/TdCoNST cla	and or 1	TAXONTS"	black	here: //	val on (TdC)	Maria	<a class="btn trans" href="#">[]</a>																		
									n trans">[Translate]																		
									pan> <a href="http://t.gg.com/p/&lt;/td&gt;&lt;/tr&gt;&lt;tr&gt;&lt;td&gt;&lt;/td&gt;&lt;td&gt;&lt;/td&gt;&lt;td&gt;&lt;/td&gt;&lt;td&gt;&lt;/td&gt;&lt;td&gt;" relay"<="" td=""><td>num-</td><td>78. &gt;xeb</td><td>ar<td>n&gt; <td></td></td></td></a>	num-	78. >xeb	ar <td>n&gt; <td></td></td>	n>  <td></td>														
	and the second sec	to the bitches would be																									
	Congrats #DeviousMaids final																										
v class="magCn	t'>squot; The wound is the place	e where light enters	yousquot;	- Rumi <a< th=""><th>href=</th><th>" class</th><th>"btn_tran</th><th>"&gt;(Trai</th><th>nslate]<div class="put&lt;/th&gt;&lt;/tr&gt;&lt;tr&gt;&lt;td&gt;class=" msgcn<br="">class="msgCn<td><pre>&gt;&gt;squot;The wound is the place t&gt;Congrats to the best cast or</pre></td><td>e where light enters ; n TV for an amazing f</td><td>you" irst seas</td><td>- Rumi<a< td=""><td>href=</td><td>class nchez 077</td><td>"btn_trans</td><td>"&gt;(Tran</td><td>nslate]</td></a<></td></div><div class="pul&lt;br&gt;aniaJRamirez @edyganem @itisiju&lt;/td&gt;&lt;/tr&gt;&lt;tr&gt;&lt;td&gt;/ class=" msgcm<br="">/ class="msgCm / class="msgCm<td>t &gt;squot; The wound is the place &gt;Congrats to the best cast or &gt;Any shockers tonight from #1</td><td>e where light enters ; n TV for an amazing f DevicusNaids finale?!</td><td>you" irst seas 11<a href<="" td=""><td>- Rumi<a< td=""><td>href= elyn_Sa</td><td>class nchez 011 trans"&gt;</td><td>"btn_trans eRealAna0: [Translate]</td><td>"&gt;[Trai</td><td>nslate)</td></a<></td></a></td></div><div c_tx5"="" class="pul&lt;br&gt;aniaJRamirez @edyganem @itisiju&lt;br&gt;div&gt;&lt;div class=" pubinfo=""></div></th></a<>	href=	" class	"btn_tran	">(Trai	nslate] <div class="put&lt;/th&gt;&lt;/tr&gt;&lt;tr&gt;&lt;td&gt;class=" msgcn<br="">class="msgCn<td><pre>&gt;&gt;squot;The wound is the place t&gt;Congrats to the best cast or</pre></td><td>e where light enters ; n TV for an amazing f</td><td>you" irst seas</td><td>- Rumi<a< td=""><td>href=</td><td>class nchez 077</td><td>"btn_trans</td><td>"&gt;(Tran</td><td>nslate]</td></a<></td></div> <div class="pul&lt;br&gt;aniaJRamirez @edyganem @itisiju&lt;/td&gt;&lt;/tr&gt;&lt;tr&gt;&lt;td&gt;/ class=" msgcm<br="">/ class="msgCm / class="msgCm<td>t &gt;squot; The wound is the place &gt;Congrats to the best cast or &gt;Any shockers tonight from #1</td><td>e where light enters ; n TV for an amazing f DevicusNaids finale?!</td><td>you" irst seas 11<a href<="" td=""><td>- Rumi<a< td=""><td>href= elyn_Sa</td><td>class nchez 011 trans"&gt;</td><td>"btn_trans eRealAna0: [Translate]</td><td>"&gt;[Trai</td><td>nslate)</td></a<></td></a></td></div> <div c_tx5"="" class="pul&lt;br&gt;aniaJRamirez @edyganem @itisiju&lt;br&gt;div&gt;&lt;div class=" pubinfo=""></div>	<pre>&gt;&gt;squot;The wound is the place t&gt;Congrats to the best cast or</pre>	e where light enters ; n TV for an amazing f	you" irst seas	- Rumi <a< td=""><td>href=</td><td>class nchez 077</td><td>"btn_trans</td><td>"&gt;(Tran</td><td>nslate]</td></a<>	href=	class nchez 077	"btn_trans	">(Tran	nslate]	t >squot; The wound is the place >Congrats to the best cast or >Any shockers tonight from #1	e where light enters ; n TV for an amazing f DevicusNaids finale?!	you" irst seas 11 <a href<="" td=""><td>- Rumi<a< td=""><td>href= elyn_Sa</td><td>class nchez 011 trans"&gt;</td><td>"btn_trans eRealAna0: [Translate]</td><td>"&gt;[Trai</td><td>nslate)</td></a<></td></a>	- Rumi <a< td=""><td>href= elyn_Sa</td><td>class nchez 011 trans"&gt;</td><td>"btn_trans eRealAna0: [Translate]</td><td>"&gt;[Trai</td><td>nslate)</td></a<>	href= elyn_Sa	class nchez 011 trans">	"btn_trans eRealAna0: [Translate]	">[Trai	nslate)
/ class="msgCn / class="msgCn / class="msgCn / class="msgCn	t >iquot;The wound is the place >Congrats to the best cast or >Any shockers tonight from #1 >Who is watching the <a href="https://www.shockers.com">https://www.shockers.com</a>	e where light enters ; n TV for an amazing f: DevicusMaids finale?! ="http://k.t.qq.com/k	you" irst seas !! <a href<br="">/DeviousM</a>	- Rumi <a on! @Ros ="#" cla aids%20m</a 	href=""	trans">	"btn_trans eRealAnsO: [Translate] 20Everyons	<pre>&gt;(Training and a second s</pre>	nslate) <div c_tx5"="" class="pul&lt;br&gt;aniaJRamirez @edyganem @itisiju&lt;br&gt;div&gt;&lt;div class=" pubinfo=""> \$20cat"&gt;#DevicusMaids marathon?</div>																		
v class="msgCn v class="msgCn v class="msgCn v class="msgCn v class="msgCn	<pre>&gt;&gt;iquot;The wound is the place &gt;&gt;Congrats to the best cast on t&gt;&gt;Any shockers tonight from #! &gt;&gt;Who is watching the <a href<br="">&gt;&gt;Beautiful day with my sister</a></pre>	<pre>e where light enters ; n TV for an amazing f: DevicusMaids finale?! ="http://k.t.qq.com/k, r Liza! Nice being how</pre>	yousquot; irst seas [! <a href:<br="">/DevicusMine<a h<="" td=""><td>- Rumi<a on! @Ros ="#" cla aids%20m ref="#"</a </td><td>href=""</td><td>trans 1253F1219</td><td>"btn_trans eRealAna0: [Translate] 20Everyone "&gt;[Translate]</td><td><pre>&gt;[Train rtiz 0D  </pre></td></a>%20can %20can %20can %20can</a>	- Rumi <a on! @Ros ="#" cla aids%20m ref="#"</a 	href=""	trans 1253F1219	"btn_trans eRealAna0: [Translate] 20Everyone ">[Translate]	<pre>&gt;[Train rtiz 0D  </pre>	nslate] <div c_tx8"="" class="pul&lt;br&gt;ania7Ramirez @edyganem @itisiju&lt;br&gt;div&gt;&lt;div class=" publnfo="">&gt; \$20cat"&gt;EDviousMaids marathon? &gt;</div> <div class="mediaWrap">&lt;</div>																		
/ class="msgCm / class="msgCm / class="msgCm / class="msgCm / class="msgCm / class="msgCm / class="fu	<pre>&gt;&gt;quot;The wound is the place t&gt;Congrats to the best cast of &gt;Any shockers toright from #} t&gt;Who is watching the <a href<br="">t&gt;Beautiful day with my siste: next&gt;<a href="http://t.qq.com">http://t.qq.com</a></a></pre>	<pre>s where light enters ; n TV for an amazing f: DeviousMaids finale?! ""http://k.t.qq.com/k r Liza! Nice being ho: /p/t/324844012849693"</pre>	youiquot; irst seas !! <a href:<br="">/DevicusMi ne<a h:<br="">class="re</a></a>	- Rumi <a on! @Ros ="#" cla aids%20m ref="#" elay" nu</a 	href=""""""""""""""""""""""""""""""""""""	class nchez 072 trans"> 1253F1210 btn_trans >Repost<,	"btn_trans eRealAna0: (Translate) 20Everyond ">(Translate) (a> <span>)</span>	<pre>&gt;[Train rtiz @Do  <!--/<br-->%20can ate]-</pre>	nslate]div>div class="pul ania7Ramicez @edyganem @itisju div>div class="publific c_tu5"> \$20cat">\$DevicusMaids marathon? >div class="mediaWrap">< div class="mediaWrap"><																		
/ class="msgCm / class="msgCm / class="msgCm / class="msgCm / class="msgCm / class="msgCm / class="fu	<pre>&gt;&gt;quot;The wound is the place t&gt;Congrats to the best cast of &gt;Any shockers toright from #} t&gt;Who is watching the <a href<br="">t&gt;Beautiful day with my siste: next&gt;<a href="http://t.qq.com">http://t.qq.com</a></a></pre>	<pre>s where light enters ; n TV for an amazing f: DeviousMaids finale?! ""http://k.t.qq.com/k r Liza! Nice being ho: /p/t/324844012849693"</pre>	youiquot; irst seas !! <a href:<br="">/DevicusMi ne<a h:<br="">class="re</a></a>	- Rumi <a on! @Ros ="#" cla aids%20m ref="#" elay" nu</a 	href=""""""""""""""""""""""""""""""""""""	class nchez 072 trans"> 1253F1210 btn_trans >Repost<,	"btn_trans eRealAna0: (Translate) 20Everyond ">(Translate) (a> <span>)</span>	<pre>&gt;[Train rtiz @Do  <!--/<br-->%20can ate]-</pre>	nslate] <div c_tx8"="" class="pul&lt;br&gt;ania7Ramirez @edyganem @itisiju&lt;br&gt;div&gt;&lt;div class=" publnfo="">&gt; \$20cat"&gt;EDviousMaids marathon? &gt;</div> <div class="mediaWrap">&lt;</div>																		
v class="magCm v class="magCm v class="magCm v class="magCm v class="magCm v class="fu v class="fu v class="magCm	<pre>t &gt;iquot:The wound is the place t &gt;Congrats to the best cast of t &gt;Any shockers tonight from # t &gt;Who is watching the <a hrefv<br="">t &gt;Beautiful day with my sister n ox &gt;<a <="" href="http://t.qq.com,&lt;br&gt;t &gt;Season Finale of #DeviousMa" pre=""></a></a></pre>	<pre>where light enters ; n TV for an amazing f: DeviousMaids finale?! ="http://k.t.qq.com/k, r Liza! Nice being hot (p/r/32484401284663" ids tonight! You can'</pre>	youiquot; irst seas !! <a href:<br="">/DevicusM me<a h:<br="">class="re t miss the</a></a>	- Rumi <a on! @Ros ="#" cla aids%20m ref="#" elay" nu e last 2</a 	href=""" href=""" """""""""""""""""""""""""""""""""	Class nchez 071 trans"> 2537421 >Con_trans >Repost	"btn_trans heRealAna00 (Translate) 20Everyond ">(Translate) (a> <span> - ise surpris</span>	<pre>&gt;&gt;(Transition of the second seco</pre>	nslate]div>div class="pul ania7Ramicez @edyganem @itisju div>div class="publific c_tu5"> \$20cat">\$DevicusMaids marathon? >div class="mediaWrap">< div class="mediaWrap"><																		
v class="magCn v class="magCn v class="magCn v class="magCn class="magCn class="magCn v class="magCn v class="magCn v class="magCn	<pre>t&gt;iquot;The wound is the place t&gt;Congrats to the best cast or t&gt;Any shockers tonight from #1 t&gt;Who is watching the <a href<br="">t&gt;Beautiful day with my sisten non"&gt;<a href="http://t.qc.com&lt;br&gt;t&gt;Beautiful day with my sisten&lt;br&gt;non"><a href="http://t.qc.com&lt;br&gt;t&gt;Beautiful day with my sisten&lt;br&gt;non"><a href="http://t.qc.com&lt;br&gt;t&gt;Beautiful day with my sisten&lt;br&gt;t&gt;Beautiful day with my sisten&lt;/td&gt;&lt;td&gt;&lt;pre&gt;s where light enters ;&lt;br&gt;n TV for an amazing f&lt;br&gt;DeviousMaids finale?!&lt;br&gt;" http:="" k,<br="" k.t.qq.com="">r Liza! Nice being ho: /p/r/324846012848653" ids tonight! You can'' king on his radio shor</a></a></a></a></pre>	youiquot; irst seas !! <a href:<br="">/DevicusM me<a h:<br="">class="re t miss the w!<a href:<="" td=""><td>- Rumi<a on! @Ros ="#" cla aids\$20m ref="#" elay" nu e last 2 ="#" cla</a </td><td>href= 1 selyn_S.s arathon class= 2 m="28" : minutes ss="bts</td><td>Class nchez 010 trans"&gt; 2537521 btn_trans &gt;Repost&lt;, s! Surpri trans"&gt;</td><td>"btn_trans eRealAnsO (Translate) 20Everyond "&gt;(Translate) (a&gt;<span> - ise surpris (Translate)</span></td><td><pre>"&gt;[Trai rtir 0D  </pre></td></a><!--/a<br-->\$20can ate]\$/span&gt;- se!!!!</a></a>	- Rumi <a on! @Ros ="#" cla aids\$20m ref="#" elay" nu e last 2 ="#" cla</a 	href= 1 selyn_S.s arathon class= 2 m="28" : minutes ss="bts	Class nchez 010 trans"> 2537521 btn_trans >Repost<, s! Surpri trans">	"btn_trans eRealAnsO (Translate) 20Everyond ">(Translate) (a> <span> - ise surpris (Translate)</span>	<pre>"&gt;[Trai rtir 0D  </pre>	nslate] <div c_ts5"="" class="pul&lt;br&gt;aniaJ2mires @edyganem @itisiju&lt;br&gt;div&gt;div class=" publnfo=""> %10cat"&gt;#DevicusMaids marathon? &gt;</div> <div class="mediaWizg">&lt; <a class="btm_trans" href="http://t.qc.com/p/t/32&lt;br&gt;a href=">[]</a></div>																			
v class="magCn v class="magCn v class="magCn v class="magCn div class="magCn v class="magCn v class="magCn v class="magCn v class="magCn	<pre>tp&gt;cqupc:The wound is the place tp&gt;cqupcast to the base cast or tp&gt;chargest to the base cast or tp&gt;chargest to the base cast or tp&gt;chargest to tp&gt;cqupcast to tp&gt;cqupcast tp&gt;cqupcast tp&gt;cqupcast tp&gt;cqupcast tp&gt;cqupcast tp&gt;cqupcast tp&gt;cqupcast tp&gt;cqupcast tp&gt;cqupcast tp&gt;cqupcast tp&gt;cqupcast tp&gt;cqupcast tp&gt;cqupcast tp&gt;cqupcast</pre>	<pre>s where light enters : n TV for an amazing f DeviousMaids finale?! ="http://k.t.qq.com/k, r Liza! Nice being ho: op/x/3242401284653" ids tonight! You can' king on his radio sho: com/p/x/3563140533534</pre>	youiquot; irst seas li <a href:<br="">/DevicusM. me<a h:<br="">class="re t miss the wi<a href:<br="">68" class</a></a></a>	- Rumi <a on! @Ros ="#" cla aids%20m ref="#" elay" nu e last 2 ="#" cla ="relay"</a 	href= 1 melyn_Sa arathon class= 1 m=28 : minuten ss="bto num="s	<pre>class nchez (TT trans"&gt; %263F%214 btn_trans &gt;Repost&lt;, s! Surpr: _trans"&gt; 24" &gt;Rep</pre>	"btn_trans heRealAnsO: (Translate) 20Everyons ">(Translate) (a> <span> - (se surpris (Translate) sst<sp< td=""><td><pre>"&gt;[Train rtiz @Di  <!--/a<br-->#20can ate]ate]sel!!!</pre></td><td>nslate]v class="magCn v class="magCn v class="magCn v class="magCn div class="magCn v class="magCn v class="magCn v class="magCn</td><td><pre>ty-iquociThe wound is the place by-Congrats to the best cast on ty-Any shockers tonight from if ty-Who is watching the <a href<br="">ty-Beautiful day with wy sisce now-&gt;ca href="http://t.qq.com ty-Season Finale of fbeviousMa ty-Season Finale of fbeviousMa ty-Ma and @MarioLoperEntra tall "unBox"&gt;ca href="http://t.qq.com ty-Season Finale of fbeviousMa ty-Season Finale of fbeviousMa ty-Season Finale of fbeviousMa ty-Season Finale of fbeviousMa ty-Season Finale of fbeviousMa ty-Shy boother @MarioLoperEntra</a></pre></td><td><pre>s where light enters ; n TV for an amazing f Devices#Asis finale?! ="http://k.t.qq.com/k r Liza! Nice being hor /p/t/324844012249693" ids tonight! You can" king on his radio shor com/p/t/3563140533534 making me laugh!da h</pre></td><td>youiquot; irst seas lika href /DevicusM. meka h class="rk t miss th wika href 60" class ref="8" class</td><td>- Rumi<a on! @Ros ="#" cla aidst20m ref="#" elay" nu e last 2 ="#" cla ="relay" lass="bt</a </td><td>href= 1 elyn_Sa ss="bts class= 1 m="28" : minut- ss="bts 'num="2 n_tran</td><td><pre>class nchez (TT trans"&gt; 253F421 btn_trans &gt;Repost&lt;, s! Surpr: trans"&gt; 24" &gt;Repo "&gt;(Trans)</pre></td><td>"btn_trans meRealAnaO: (Translate) 20Everyons "&gt;(Translate) a<span> - ise surpri: (Translate) st<sp late)&lt;</sp </span></td><td><pre>"&gt;[Trai rtiz @D  <!--/a<br-->%20can ate]%/span&gt;- se!!!!</pre> //a&gt;<!--/a<br-->an&gt;!</td><td>niltaj/(x&gt;anix7anics" delyanem dirisiyu div&gt;<div<les="public" dirisiyu<br="">v/div&gt;<div<less="public" dirisiyu<br=""><div<less="public" dirisi"<br=""><a <br="" href="tup://t.q.com/p//32&lt;br&gt;a href=" p="" t.q.com="" tup:="">div&gt;div<less="public" dirisi"<br="">div&gt;div<less="public" dirisi"<="" td=""></less="public"></less="public"></a></div<less="public"></div<less="public"></div<les="public"></td></sp<></span>	<pre>"&gt;[Train rtiz @Di  <!--/a<br-->#20can ate]ate]sel!!!</pre>	nslate]v class="magCn v class="magCn v class="magCn v class="magCn div class="magCn v class="magCn v class="magCn v class="magCn	<pre>ty-iquociThe wound is the place by-Congrats to the best cast on ty-Any shockers tonight from if ty-Who is watching the <a href<br="">ty-Beautiful day with wy sisce now-&gt;ca href="http://t.qq.com ty-Season Finale of fbeviousMa ty-Season Finale of fbeviousMa ty-Ma and @MarioLoperEntra tall "unBox"&gt;ca href="http://t.qq.com ty-Season Finale of fbeviousMa ty-Season Finale of fbeviousMa ty-Season Finale of fbeviousMa ty-Season Finale of fbeviousMa ty-Season Finale of fbeviousMa ty-Shy boother @MarioLoperEntra</a></pre>	<pre>s where light enters ; n TV for an amazing f Devices#Asis finale?! ="http://k.t.qq.com/k r Liza! Nice being hor /p/t/324844012249693" ids tonight! You can" king on his radio shor com/p/t/3563140533534 making me laugh!da h</pre>	youiquot; irst seas lika href /DevicusM. meka h class="rk t miss th wika href 60" class ref="8" class	- Rumi <a on! @Ros ="#" cla aidst20m ref="#" elay" nu e last 2 ="#" cla ="relay" lass="bt</a 	href= 1 elyn_Sa ss="bts class= 1 m="28" : minut- ss="bts 'num="2 n_tran	<pre>class nchez (TT trans"&gt; 253F421 btn_trans &gt;Repost&lt;, s! Surpr: trans"&gt; 24" &gt;Repo "&gt;(Trans)</pre>	"btn_trans meRealAnaO: (Translate) 20Everyons ">(Translate) a <span> - ise surpri: (Translate) st<sp late)&lt;</sp </span>	<pre>"&gt;[Trai rtiz @D  <!--/a<br-->%20can ate]%/span&gt;- se!!!!</pre> //a> /a<br an>!	niltaj/(x>anix7anics" delyanem dirisiyu div> <div<les="public" dirisiyu<br="">v/div&gt;<div<less="public" dirisiyu<br=""><div<less="public" dirisi"<br=""><a <br="" href="tup://t.q.com/p//32&lt;br&gt;a href=" p="" t.q.com="" tup:="">div&gt;div<less="public" dirisi"<br="">div&gt;div<less="public" dirisi"<="" td=""></less="public"></less="public"></a></div<less="public"></div<less="public"></div<les="public">									
v class="msqCn v class="msqCn v class="msqCn v class="msqCn v class="msqCn v class="msqCn v class="msqCn v class="msqCn v class="msqCn n> <div class="msqCn&lt;br&gt;n&gt;&lt;div class=" msqcn<="" td=""><td><pre>tripiquotiffe vound is the place to Congrate to the best case o to Namy shockers tonsight from if the Wind is watching the <a href<br="">to Mane is watching the <a href<br="">to Mane is a tripical of the shock to Mane is the shock of the shock of the tripical of the shock of the shock of the tripical of the shock of the sh</a></a></pre></td><td><pre>e where light enters ; n TV for an amazing f DevicusMatis finale?! ="http://k.t.qc.com/k r Lima! Nice being ho /p/r/324844012484503 ids tonight! You can' king on his radio sho com/p/r/3563140533534 making me laugh!ca h om/p/r/32632412712791</pre></td><td>youlquot; irst seas !!<a href-<br="">/DevicusM me<a h<br="">class="r t miss th v!<a href-<br="">v!<a href-<br="">v!<a href-<br="">tef="#" class=" class=" class="</a></a></a></a></a></td><td>- Rumi<a on! @Ros ="#" cla aidst20m ref="#" elay" nu e last 2 ="#" cla ="relay" lass="bt "relay"</a </td><td>href= 1 selyn_S.s marathou class= 1 m="28" minuter ss="bts num="13 num="3</td><td><pre>class nchez 0T trans"&gt; 25354210 btn_trans &gt;Repost&lt;, s! Surpr: trans"&gt; 24" &gt;Repost "&gt;(Trans) "&gt;Repost</pre></td><td>"btn_trans meRealAna0: (Translate) 20Everyond "&gt;(Translate) (a&gt;capan&gt;)- use surpri: (Translate) stcapan: (</td><td><pre>"&gt;[Trai rtiz @D  <!--/a<br-->%20can %20can ate]% sel!!!</pre> /div&gt; /div&gt; /div&gt; /span &gt; /div&gt; /span /div&gt; /span /div&gt; /span /div&gt; /span /div&gt; /span /div&gt; /span /div&gt; /span /div&gt; /span /span /div&gt; /span /s</td><td>malatai(ab-/div-div-div-div-div- diu-div-div-div-div-div-div- div-div-div-dia-s-publiche _sai- \$20cat"&gt;fDevicuMaids marathon7 <dab-div-div-dis-s-madiahrag">di a hafe***_loss="madiaHrag"&gt;di div-div-dis-s-madiaHrag"&gt;di div-div-lass="madiaHrag"&gt;di div-div-lass="madiaHrag"&gt;di div-div-lass="madiaHrag"&gt;di div-div-lass="madiaHrag"&gt;di div-div-lass="madiaHrag"&gt;di div-div-lass="madiaHrag"&gt;di div-div-lass="madiaHrag"&gt;di div-div-lass="madiaHrag"&gt;di div-div-div-lass="madiaHrag"&gt;di div-div-div-lass="madiaHrag"&gt;di div-div-div-div-lass="madiaHrag"&gt;di div-div-div-div-div-div-div-div-div-div-</dab-div-div-dis-s-madiahrag"></td></div>	<pre>tripiquotiffe vound is the place to Congrate to the best case o to Namy shockers tonsight from if the Wind is watching the <a href<br="">to Mane is watching the <a href<br="">to Mane is a tripical of the shock to Mane is the shock of the shock of the tripical of the shock of the shock of the tripical of the shock of the sh</a></a></pre>	<pre>e where light enters ; n TV for an amazing f DevicusMatis finale?! ="http://k.t.qc.com/k r Lima! Nice being ho /p/r/324844012484503 ids tonight! You can' king on his radio sho com/p/r/3563140533534 making me laugh!ca h om/p/r/32632412712791</pre>	youlquot; irst seas !! <a href-<br="">/DevicusM me<a h<br="">class="r t miss th v!<a href-<br="">v!<a href-<br="">v!<a href-<br="">tef="#" class=" class=" class="</a></a></a></a></a>	- Rumi <a on! @Ros ="#" cla aidst20m ref="#" elay" nu e last 2 ="#" cla ="relay" lass="bt "relay"</a 	href= 1 selyn_S.s marathou class= 1 m="28" minuter ss="bts num="13 num="3	<pre>class nchez 0T trans"&gt; 25354210 btn_trans &gt;Repost&lt;, s! Surpr: trans"&gt; 24" &gt;Repost "&gt;(Trans) "&gt;Repost</pre>	"btn_trans meRealAna0: (Translate) 20Everyond ">(Translate) (a>capan>)- use surpri: (Translate) stcapan: (	<pre>"&gt;[Trai rtiz @D  <!--/a<br-->%20can %20can ate]% sel!!!</pre> /div> /div> /div> /span > /div> /span /div> /span /div> /span /div> /span /div> /span /div> /span /div> /span /div> /span /span /div> /span /s	malatai(ab-/div-div-div-div-div- diu-div-div-div-div-div-div- div-div-div-dia-s-publiche _sai- \$20cat">fDevicuMaids marathon7 <dab-div-div-dis-s-madiahrag">di a hafe***_loss="madiaHrag"&gt;di div-div-dis-s-madiaHrag"&gt;di div-div-lass="madiaHrag"&gt;di div-div-lass="madiaHrag"&gt;di div-div-lass="madiaHrag"&gt;di div-div-lass="madiaHrag"&gt;di div-div-lass="madiaHrag"&gt;di div-div-lass="madiaHrag"&gt;di div-div-lass="madiaHrag"&gt;di div-div-lass="madiaHrag"&gt;di div-div-div-lass="madiaHrag"&gt;di div-div-div-lass="madiaHrag"&gt;di div-div-div-div-lass="madiaHrag"&gt;di div-div-div-div-div-div-div-div-div-div-</dab-div-div-dis-s-madiahrag">																		
v class="magCn v class="magCn v class="magCn v class="magCn v class="magCn v class="magCn v class="magCn an> <div class="magCn&lt;br&gt;n&gt;&lt;div class=" magcn<br="">v class="magCn</div>	representation of the place	<pre>s where light enters in TV for an amazing f DeviousKaids finale1: "http://k.t.qt.cqt.com/k r Liza!Nice being hot /pt/134444012434953" ids tonight!You can' king on his radio sho com/pt/13453140533534 making me laugh!<a 2428341271287<br="" c="" h="" homp="">ay on Extra k tea here</a></pre>	youlquot; irst seas !! <a href-<br="">/DevicusM me<a h<br="">class="r t miss the v!<a href-<br="">60" class ref="8" class= s" class=</a></a></a>	- Rumi <a on! @Ros ="#" cla aids%20m ref="#" elay" nu e last 2 ="#" cla "relay" "selay" "btn_tra</a 	href= 1 selyn_S.s marathou class= 2 me"28 = 1 minuter ss="bts nume"13 nume"13 nume"3 nume"3	<pre>* class nchez 0T trans"&gt; t253Ft21t btn_trans &gt;Repost&lt;, s! Surpr: trans"&gt; 24" &gt;Repo "&gt;[Trans" * Repost anslate]</pre>	"btn_trans meRealAna0; Translate 20Everyond ">[Translate a> <spa>! Ise surpri: Translate st<spa: (<a><dp>://a&gt;</dp></a></spa: </spa>	<pre>&gt;[Trai rti: 0D  ate]ate]sel!!! <!--/a<br-->an&gt;!- /div&gt;&gt;(/span&gt;- /div&gt;control (/span)</pre>	nilea(/a>/di>/di>/di>/di>/di>/di>/di>/di>/di>/di																		
v class="msGCn v class="msGCn v class="msGCn v class="msGCn v class="msGCn v class="msGCn v class="msGCn v class="msGCn v> class="msGCn v> class="msGCn v> class="msGCn div class="sgCn	<pre>n &gt;&gt;quotentine wound is the place t &gt;&gt;quotentine to the best cast of t &gt;&gt;my shockers tenight from if the place the place the place the place to =&gt;months == months == months == months the place the place the place the place under"&gt;=&gt; here"http://t.qq.om &gt;&gt;months == months == months == months under"&gt;=&gt; here"http://t.qq.om &gt;&gt;&gt;months ==&gt; months ==&gt;</pre>	<pre>s where light enters a n TV for an emaring f DeviousKaids finale?! ="http://k.t.tq.cocm/k kr Liza! Nice being hor r/p/s/32464401284653" ids tondpht I You can" king on his radio show cocm/p/s/362140531254 making me laughica h om/p/s/3628141731257 ay on Extraits heref=" r/p/s/36341031057257"</pre>	youtquot; irst seas !! <a href:<br="">/DevicusM me<a h<br="">class="r t miss the v!<a href:<br="">60" class ref="1" cl 5" class=" class="ref class="ref</a></a></a>	- Rumi <a on! @Ros ="#" cla aids%20m ref="#" elay" nu e last 2 ="#" cla ""relay" lass="bt "relay" "btn_tra elay" nu</a 	href="bi melyn_Ss marathos class="bi me"28" "nume"3 "nume"3 mume"3 ms">[To me"25"	<pre>class nchez 0T trans"&gt; t253Ft21t btn_trans &gt;Repost&lt;, s! Surpr: trans"&gt; 24" &gt;Repo "&gt;[Trans" "&gt;[Trans" "&gt;Repost anslate]</pre>	"btn_tran meRealAna0; Translate; 20Everyond; a> <span>1; se surpri; Translate; st<span; ate]&lt;; a&gt;<span; ate]&lt;; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span; a&gt;<span< td=""><td><pre>&gt;&gt;[Trained to be a constructed by a construction of the action of t</pre></td><td>malatej (ab/div/div/div/div/div/div/dis/ malat2Nanicz div/div/div/dis/edypane %itisju div/div/div/dis/markhon7 /div/div/dis/markhon7 /div/div/dis/markhon7 /div/div/dis/markhon7 /div/div/dis/markhon7 /div/div/dis/markhon7 /div/div/dis/markhon7 /div/div/div/dis/markhon7 /div/div/div/div/div/div/div/div/di</td></span<></span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span; </span>	<pre>&gt;&gt;[Trained to be a constructed by a construction of the action of t</pre>	malatej (ab/div/div/div/div/div/div/dis/ malat2Nanicz div/div/div/dis/edypane %itisju div/div/div/dis/markhon7 /div/div/dis/markhon7 /div/div/dis/markhon7 /div/div/dis/markhon7 /div/div/dis/markhon7 /div/div/dis/markhon7 /div/div/dis/markhon7 /div/div/div/dis/markhon7 /div/div/div/div/div/div/div/div/di																		
v class="msGCn v class="msGCn v class="msGCn v class="msGCn div class="su v class="msGCn an> <div class="msGCn&lt;br&gt;an&gt;&lt;div class=" msgcn<br="">v class="msGCn v class="msGCn v class="msGCn v class="msGCn v class="msGCn v class="msGCn</div>	<pre>projectifie wound is the place programs to the best cast of plany shockers tendpd from if plane the place of the place plane the place of the place place of the place of t</pre>	<pre>s where light enters; n TV for an amazing f deviculation finals?) "http://k.t.qq.com/k to lisa Nice being hot /p/r/32444012484503 ids conight: Nou can' king on his radio short om/p/r/24284412717879 yo n Extraits heref*" /p/r/3044018072874" lo of Edvicumations?</pre>	youlquot; irst seas !! <a href:<br="">/DevicusM me<a h<br="">class="r t miss th v!<a href:<br="">60" class= ref="8" c 50" class= class="r class="r Lifetime</a></a></a>	- Rumi <a on! @Ros ="#" cla aids%20m ref="#" elay" nu e last 2 ="#" cla ="relay" lass="bt "relay" btn_tra elay" nu will be</a 	href= 1 elyn_Sa arathon class= "bto ms=28" i minute ss="bto ' nume" 3 nume" 3 ms=>[To ms=>[To ms=>[To ms=25" i airing	<pre>class nchez 0T trans"&gt; 253F211 btn_trans Repost&lt;, s! Surpr: trans"&gt; 24" &gt;Repost&lt;, s! Surpr: trans"&gt; 24" &gt;Repost&lt;, a maratl</pre>	"btn_trani eRealAnaO: (Translate) 20Everyond ">(Translate) (a> <span> - (se surpri) (Translate) (<a><span> - (<a><span> - (<a><da>&gt; (a&gt;&lt;<span> - (a&gt;<span> - (a&gt;<s< td=""><td><pre>&gt;&gt;[Train rtis @D  </pre></td></s<></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></da></a> *20can ate](/span&gt;- se!!!!<!--<br--> </span></a><!--/or--> *20can ate]*/span&gt;- */spa</span></a></span>	<pre>&gt;&gt;[Train rtis @D  </pre>	nilea(/a>/(div/div/div/div/div/div/div/div/div/div/																		
v class="magCn v class="magCn v class="magCn v class="magCn v class="magCn v class="magCn v class="magCn v class="magCn n> <div class="magCn&lt;br&gt;&lt;div class=" magcn<br="">v class="magCn v class="magCn v class="magCn v class="magCn v class="magCn v class="magCn v class="magCn v class="magCn v class="magCn</div>	<pre>tripicoutile vound is the place tricogrates to the best case of tricogrates to the best case of tricogrates to the best case of tripicous tri</pre>	<pre>where light enters in TV for an amazing f Devicus/kids finale7) "http://k.cg.com/k "http://k.cg.com/k lids conight/ Tou can" king on his radio sho com/p//s/2814/0333344 making me laughta h making me laughta h ump/r/s/28141271373 ay on Extraita het6"" (p/r/s03412012718") le of Devicus/Maids11</pre>	youtquot; irst seas /DevicusM /DevicusM me <a h<br="">class="re t miss the vi<a href<br="">60" class ref="8" class=" class="re class="re class="re lifetime a href="8"</a></a>	- Rumida on! @Ros ="#" cla aidst20m ref="#" elay" cla ="#" cla ="relay" lass="bt "relay" "btn_tra elay" nu will be " class="	<pre>http://www.salaction.com/ isserverses/ classer 1 mm="28" is isserverses/ inum="14" inum="14" ins="&gt;inum="14" ins="25" is isserverses/ ins="25" is isserverses/ isserverse</pre>	<pre>class* nchez @TM trans"&gt; 253F%21 btn_trans &gt;Repost&lt;, s! Surp: trans"&gt; 24" &gt;Report &gt;[Trans] &gt;Repost&lt;, a maratl ans"&gt;[Tr</pre>	"btn_trans eRealAngO: [Translate] 20Everyond ">[Translate] se surpri: [Translate] st <span>[- (a- (a&gt;&lt;</span>	<pre>&gt;&gt;[Train rtis @Di  <!--/<br-->*20cant ate]- sel!!!&lt; (/span&gt;- /div&gt;- /div&gt;- /div&gt;- sel</pre>	niles(/a>/di>/di>/di>/di>/di>/di>/di>/di>/di>/di																		
v class="magGn v class="magGn v class="magGn v class="magGn <div class="fug&lt;br&gt;v class=" maggn<br="">an&gt;ddiv class="fug v class="magGn n&gt;<div class="magGn&lt;br&gt;v class=" maggn<br="">v class="magGn v class="ma</div></div>	p>equivilies yound is the place p>equivilies to the best cast of p>hay shockers tonight from if p>hay shockers tonight from if p>heathing to the participant p>heathing to the particip	where light enters in TV for a maxing f Devicus/isf finale7 "http://k.cg.com/k "http://k.cg.com/k final/isf ids conight! You can" /g/v/386310033154 ids conight! You can" making me laugh!ch http:// making me laugh!ch http:// ay on Extrate.http:// yo/v38041018028754 t of Devicus/aisf t chataau Marmont	youiquot; irst seas !! <a href:<br="">class="r t miss th w!<a href:<br="">60" class="s class="s class="r class="r lifetime a href="#"</a></a>	- Rumida on! @Ros =*g" cla aids420m ref="g" elay" nu e last 2 =*g" cla "relay" "btn_tra elay" nu will be " class=" class="	<pre>h href= 4 elyn_Sa sss=bta arathon class= bta i minuter sss=bta i mum="3 ins"&gt;[Trainer=25" i airine ''Dtn_t ''num="1</pre>	<pre>* class* nchez @TM trans"&gt; 2557421 0557421 05555421 05555421 05555421 05555421 05555421 0555542100000000000000000000000000000000</pre>	"btn_trans eRealAnaO: [Translate] 20Everyone ">[Translate] st <span> - [Translate] st<span> - (a&gt;<span> - (a&gt;)</span> - (a&gt;<span> - (a&gt;)</span> -</span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span></span>	<pre>&gt;&gt;Train 00  <!--/ --></pre>	nilea(/a>/(div/div/div/div/div/div/div/div/div/div/																		

Fig. 9.9 The crawled content

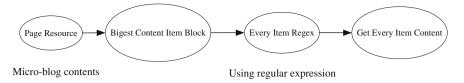


Fig. 9.10 The processing flow

```
// Get the content of corresponding Urls
Algorithm 3 (i.e., grabOnePageArticle(String url, boolean
isFirst))
Input:
(1) Url //pages to be crawled;
(2) Url_Refer //one of the header parameter within the HTTP
request;
(3) Url_Host //one of the header parameter such the port
number, etc.
Output:
(1) content.
Step1. Url=getCompleteUrl(time,currentPage,aid,uid) // Fill in
the Url parameters dynamically, including the current time,
currentPage, aid parameter (i.e., id parameters of the micro-
blog contents, see Fig.\,\ref{fig:9}), uid parameter
(i.e., user name).
Step2. content=grabPageSource4QQ.grabPageSource0fQQ(Url,
Url_Refer, Url_Host);
```

The data parsing module needs to process the crawled data and parse the microblog content. The crawled data can be classified into two corpus, i.e., plaintext and cipher text. As the plaintext is regular and uniform, we use the regular expression to extract the real contents (e.g., the micro-blog contents, Url, id, published time, IP address, reviewed or commented number, forwarded number, etc.). The processing flow is shown in Fig. 9.10.

## 9.6 Experimental Results and Analysis

## 9.6.1 About the Testing Data Set and the Experimental Environment

As for these micro-blog big data, its persistence is an important issue. We use the *MongoDB*, *Redis* to store and cache them in our real application, and *Mysql* is

(a)			(b)			
wi	name	sex	sendUrl	docUrl	article	person_id
http://Lqq.com/EvaLongoriavip	EvaLongoria	femail	http://Lgg.com/EvaLongoriavip	http://t.gq.com/p/t/163872010546035	Just amazing! http://url.cn/TdCoN8[Translate]	EvaLongoriavip
http://Lqq.com/BeverleyM	Beverley	femail	http://Lgg.com/Evalongoriavip	http://Lgq.com/p/t/342346059777923	Fun day with the nieces and nephews! Time to make their Hallo	
http://Lqq.com/PaulScheer	PaulScheer	mail				
http://Lqq.com/neilgaimanvip	NeilGaiman	mail	http://Lqq.com/EvaLongoriavip	http://Lqq.com/p/t/339346042208598	Congrats #DeviousMaids finale was highest rated telecast of e	
http://Lgq.com/JohnStamos	JohnStamos	mail	http://t.qq.com/PaulScheer	http://t.qq.com/p/t/348849124784899	.@AnthonySHead is a very bad man who hates San Diego on the	PaulScheer
http://et.qq.com/KaylaCollins	KaylaCollins	femail	http://Lqq.com/PaulScheer	http://Lqq.com/p/t/337348059180032	My wife with @MsJuneDiane (New Girl, Flight of the Conchords	PaulScheer
http://l.qq.com/SnoopDogg	SnoopDogg	mail	http://Lqq.com/PaulScheer	http://t.qq.com/p/t/333848030426112	Meeting an important Hollywood friend to discuss a project.[To	PaulScheer
http://Lqq.com/Nacho_Polo	Nacho_Polo	mail	http://Lqq.com/PaulScheer	http://t.qq.com/p/t/310387092696122	NTSESDSUV: is going to London (for Reall) for our TV Movie.	PaulScheer
http://Lqq.com/JeffGiburg	JeffGiburg	mail	http://Log.com/OrlandoJones	http://t.gg.com/p/t/34997129217626	Huh_ well_ The more you know. :/ http://url.cn/KS3Cit[Transla	
http://Lqq.com/CarolynHennesy	CarolynHennesy	femail	http://Log.com/OrlandoJones	http://t.qq.com/p/t/258643107676980	Seriously Congress can't you get your freaking priorities strai	
http://Lqq.com/NikkiReed	NikkiReed	femail				
http://Lqq.com/AbigaiSpencer	AbigailSpencer	femail	http://Lqq.com/AliciaWitt	http://Lqq.com/p/t/227077002802688	10 minutes to showtime! http://url.cn/P9Tau3[Translate]	AliciaWitt
http://Lqq.com/Bonnie/ilLafin	BonnieJillLafin	femail	http://t.qq.com/DarrenRovell	http://t.qq.com/p/t/323345120290114	Nov. 13, 1994: Last time the Bengals, Browns and Lions won o	DarrenRovell
http://Lgg.com/LukeBilyk	LukeBilyk	mail	http://Lqq.com/DarrenRovell	http://Lqq.com/p/t/321342042720847	Today's payouts: Northwestern paying Maine \$450K, Nebraska	DarrenRovell
http://Lqq.com/JamesPhelan	JamesPhelan	femail	http://t.gg.com/cocorochaonline	http://t.qq.com/p/t/329349120476115	New cover for all my Chinese friends!!! Harpers Bazaar China -	cocorochaonline
http://Lqq.com/winds_official	winds_official	mail	http://t.og.com/cocorochaonline	http://t.gq.com/p/t/339348036456824	Wilhelmina Models takes you behind-the-scenes on Coco Roc	cocorochaonline
http://Lqq.com/PeytonList	PeytonList	femail			Watch with Allure Magazine as I chop off my long hair into a p	
http://Lqq.com/jamieloxwip	jamiefoxxvip	mail		http://Lgq.com/p/t/276017069296076		
http://t.on.com/Carrieleninaha	CarrieAnelnaha	femail	http://Lqq.com/cocorochaonline	http://t.qq.com/p/t/243927112474143	So excited to be at the #MetropolitanOpera for opening night	cocorochaonline

Fig. 9.11 The parsed results, a account information b other parsed content

only used as the experimental platform. Figure 9.11 shows the *MySql*-based parsed Tecent\_Micro-blog data.

In order to evaluate the algorithm's performance, we classify the following tested data set according to their authorities into three classes. The reason we use the following three classes is that the micro-blog official platform usually presents the following three different classes, and the three classes have different data sizes. In detail, the first class has less data while the two others have more. In detail, the first class is the ordinary (i.e., minor authorities) microbloggers, whose propagative scope is only limited within old friends or classmates; the second class is those medium authorities' microbloggers (e.g., network magazines microblog), whose contents can be followed or spread into all kinds of users; the last class is famous persons' microblogs. We will test the performance on the above different data environment.

## 9.6.2 Ordinary Microblogger's Performance Evaluation

In this section, we use someone's microblog as an example. Figure 9.12a shows the original microblog, and b shows the crawled and parsed data, respectively. It is clear that there is no difference between them, and all contents have been obtained correctly.

#### 9.6.3 Medium Authorities' Microblogger Performance Evaluation

Here, we use some medium authorities' microbloggers as the experimental platform. Figure 9.13a shows the micro-blog interface, while b shows the crawled and parsed data.

9 Applied Methods and Techniques for Modeling and Control

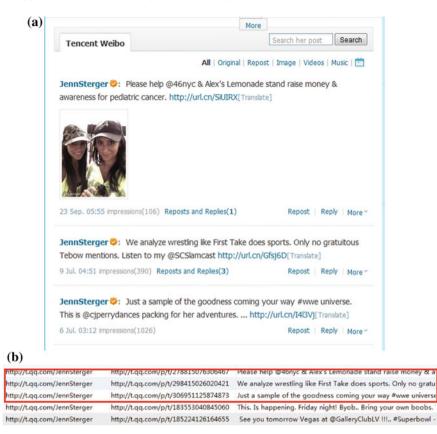


Fig. 9.12 Ordinary and the parsed results of the ordinary microblogger's, **a** ordinary data, **b** parsed results

# 9.6.4 Famous Persons' Microblog

As for these famous persons' microblogs, we use the famous actor Tom Cruise's microblog as an example. Figure 9.14a shows his micro-blog interface, while b shows the crawled and parsed data.

## 9.6.5 Performance Evaluation

Sometimes micro-blog services provide some APIs. Through these services, the well-structured data can be easily obtained, so it can provide us the probability of constructing uniform and universal software architecture to utilize the provided APIs to automatically download data. However, there are usually some limits and

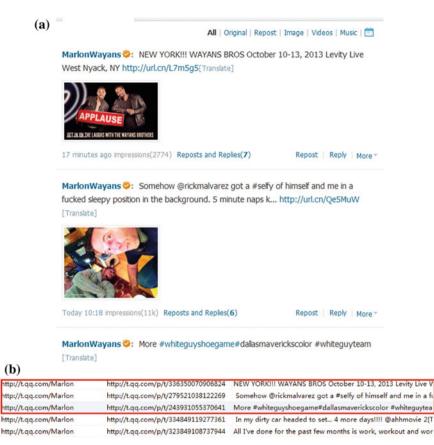


Fig. 9.13 Medium authorities' microblogger and the corresponding parsed result, **a** ordinary data, **b** parsed results

obstacles. In order to evaluate the performance, we present the comparison between APIs-based crawling and the simulating browser behavior approach, see Table 9.1.

From the above comparison, as for the proposed method, it is clear that the parsed data's degree of integrity and the accuracy or scope is higher. But, as shown before, if the template or the main framework of the microblog has been changed, the accuracy of the parsed data is lower than usual. Fortunately, these changes occur rarely. If we can track or analyze the parsed data periodically, it is easy to find the changes and then revise some special rules to parse the corresponding data.

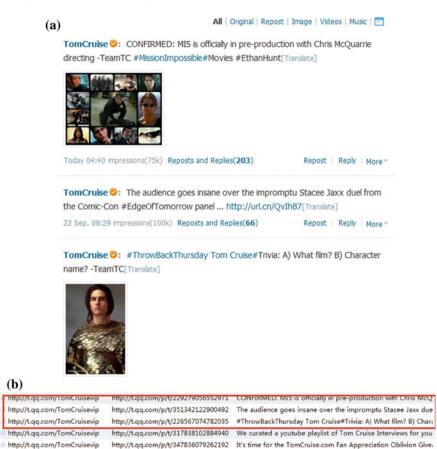


Fig. 9.14 Famous persons' micro-blog and the parsed results, a ordinary data, b parsed results

# 9.7 Conclusion

It is hard for a traditional web page crawler to crawl micro-blog data as usual, and most microblogs' official platforms cannot offer some suitable tools or RPC interfaces to collect the data effectively and efficiently. This chapter presents some algorithms and strategies on crawling and parsing micro-blog data effectively based on simulating browsers' behaviors. This needs to analyze the simulated browsing behavior in order to obtain the requesting URLs, to simulate and analyze the sending URLs requests according to the order of data sequence. It needs to focus on crawling on some special crowds and crawl some special contents by using the microblog's searching function. Parallel crawling and the multiprocessing technology are also used to download the data simultaneously. The experimental results and the analysis show the feasibility of the approach. Existing works are also presented at the end.

Performance	Algorithms	
	The proposed simulating browser behavior-based crawling	Official APIs-based crawling
About the micro-blog data entrance	Micro-blog platform account	<ol> <li>Micro-blog platform account</li> <li>AppKey</li> <li>AccessToken</li> </ol>
About the data usability	Flexible and easy to use	<ul> <li>(1) Existing more</li> <li>limitations such as fixed</li> <li>or lower crawling</li> <li>frequency, it is not</li> <li>flexible</li> </ul>
		(2) The parsed contents are usually fixed and researchers cannot choose some special categories from them
About the multi-threads performance	Better	Poor
About the crawling frequency	Flexible, and the only limitation is the bandwidth	Not flexible, as there are more limitations such as API invoking rights or authorization
About the stability	Better. The only limitation is the variation of the Official's platform, but the variation is rare	Better
About the data integrity	Better, and almost all needed data can be parsed	Poor, and some data cannot be parsed such as the <i>docUrl</i> , etc. <i>DocUrl</i> parameter can be seen from the Fig. 9.12b
About visiting or data requested frequency	Larger than 60 times (per minutes)	Less than 1 times (per minutes)
About the degree of data integrity	100 %	95 %, lacking some materials such as <i>docUrl</i> , etc
About the data accuracy	99 %	100 %
About the data scope	Widely, and almost all the data appearing on the micro-blog platform can be crawled and parsed	Narrowly, and the only small new parts, such as microbar, microgroup, etc., can be obtained

 Table 9.1
 The performance comparison of the two approaches

**Acknowledgments** Some earlier works were done in Beijing Institute of Technology with the help of Dr. Hua-ping Zhang and Prof. Yin-ping Zhao. This work is sponsored by the National Science Foundation of Hebei Province (No. F2013208105) and the National Science Foundation of China (No. 61272362). It is also sponsored by Hebei Province Scientific and Technical Key Task (No. 12213516D).

## References

- Kwak H, Lee C, Park H, Moon S (2010) What is Twitter, a social network or a news media? In: 19th international conference on world wide web. ACM Press, USA, pp 591–600
- Weng J, Lim EP, Jiang J, He Q (2010) TwitterRank: finding topic-sensitive influential twitterers. In: 3rd international conference on web search and web data mining. ACM Press, USA, pp 261–270
- Cristian DNM, Lee L, Bo P, Kleinberg J (2012) Echoes of power: language effects and power differences in social interaction. In: 21th international conference on world wide web. ACM Press, France, pp 699–708
- 4. Wu S, Hofman JM, Mason WA, Watts DJ (2011) Who says what to whom on Twitter. In: 20th international conference on the world wide web. ACM Press, India, pp 705–714
- Abel F, Gao Q, Houben GJ, Tao K (2011) Analyzing user modeling on Twitter for personalized news recommendations. In: International conference on user modeling, adaptation and personalization. LNCS, vol 6787. Springer, Spain, pp 1–12
- Chen J, Nairn R, Nelson L, Bernstein M, Chi E (2010) Short and tweet: experiments on recommending content from information streams. In: 28th international conference on human factors in computing systems. ACM Press, USA, pp 1185–1194
- Bakshy E, Hofman JM, Mason WA, Watts DJ (2011) Everyone's an influencer: quantifying influence on Twitter. In: 3rd international conference on web search and data mining. ACM Press, Hong Kong, pp 65–74
- Bakshy E, Rosenn I, Marlow C, Marlow C (2012) The role of social networks in information diffusion. In: International conference on world wide web. ACM Press, France, pp 519–528
- Sachan M, Contractor D, Tanveer AF, Subramaniam LV (2012) Using content and interactions for discovering communities in social networks. In: International conference on world wide web. ACM Press, France, pp 331–340
- Dan C, Shipman FM (2009) Capturing on-line social network link dynamics using eventdriven sampling. In: International conference on computational science and engineering, vol 4. Vancouver, Canada, pp 284–291
- Goyal A, Bonchi F, Lakshmanan LV (2010) Learning influence probabilities in social networks. In: 3th international conference on web search and data mining. ACM Press, USA, pp 241–250
- Agarwal A, Durgesh S, Pandey AKA, Goel V (2012) Design of a parallel migrating web crawler. J Adv Res Comput Sci Softw Eng 2(4):147–153
- Kim KS, Kim KY, Lee KH, Kim TK, Cho WS (2012) Design and implementation of web crawler based on dynamic web collection cycle. In: International conference on information networking (ICOIN). Bali, Indonesia, pp 562–566
- Chandramouli A, Gauch S, Eno J (2012) A cooperative approach to web crawler URL ordering, human–computer systems interaction: backgrounds and applications. J Adv Intell Soft Comput 98:343–357
- Lu G, Liu S, Lü K (2013) MBCrawler: a software architecture for micro-blog crawler. In: International conference on information technology and software engineering. Lecture Notes in Electrical Engineering, vol 212. Springer, Berlin, Heidelberg, pp 119–127
- Gao K, Li SW (2010) The cooperation model for multi agents and the identification on replicated collections for web crawler. Int J Model Identif Control 11(3–4):224–231

- Garg A, Tai K (2013) Comparison of statistical and machine learning methods in modelling of data with multicollinearity. Int J Model Identif Control 18(4):295–312
- Han G, Zhu H, Ge J (2013) Effective search space reduction for human pose estimation with Viterbi recurrence algorithm. Int J Model Identif Control 18(4):341–348
- Singh S, Mittal P, Kahlon KS (2013) Empirical model for predicting high, medium and low severity faults using object oriented metrics in Mozilla Firefox. Int J Comput Appl Technol 47(2/3):110–124
- 20. HttpWatch: Introduction to HttpWatch 8.x (2013). http://help.httpwatch.com/#introduction. html
- 21. Ajax: Introduction to Ajax (2013). http://api.jquery.com/category/ajax/
- 22. Json: Introduction to Json (2013). http://www.json.org/index.html