Extracting Web Data Connections for Identifying Similar Interests in Microblogging

Samantha Vrabl [1], Jonice Oliveira[1,2], Cláudia L. R. Motta[1,3]

[1]Graduation Program in Computer Science (PPGI) [2] Computing Science Department (DCC) / Institute of Mathematics (IM) [3] Nucleo of Electronic Computer (NCE) - Federal University of Rio de Janeiro - Rio de Janeiro - RJ, Brazil

svrabl@hotmail.com, jonice@dcc.ufrj.br, claudiam@nce.ufrj.br

Abstract. *Share experiences in Web rely on connections and data exchange, while social networks provide means of update about an interest through people's lenses. Our knowledge acquisition process can be an optimized and a nice experience when we meet other's who share same interests. This paper describes a social match model that focus in managing web data connections in a microblogging. Based on Twitter's data, we developed level of knowledge indicators and identified profile traces, which together, can offer a more precise people recommendation.*

1. Introduction

Our research goal is to offer optimization in knowledge acquisition process by approaching people with similar interests. Giving the information from one person to another who shares particular similar interest, offering new trends of data search and optimizing the learning curve, since it shows themes that have been already explored instead of starting a search from scratch. As consequence we also have a strengthening of social network, when it's known that the new information is going to be shared with, collaborative learning is enabled.

In Section 2, we describe the reasons we decided to adopt microblogging as the environment. We describe it and detail its usage dynamic, and also presents the related bibliography. Section 3 explicates the social match model proposed and implementation details. In Section 4, we explore related works and the research relevance. Finally we pointed out some conclusions and future proceedings.

2. Microblogging

Research about Microblogging has been increased, since 2007, one year after the most popular service Twitter has been created. Several papers started to evaluate its messages content and user's intention and behavior (Java et al, 2007) (Krishnamurthy et al, 2008) (Miller, 2008) (Voida et al, 2008). It was also discussed in practices such as mobile-georeference systems (Barkhuus et al, 2008); semantic and distributed approach (Passant et al, 2008) and social media usage for disaster responses (Sutton et al, 2008).

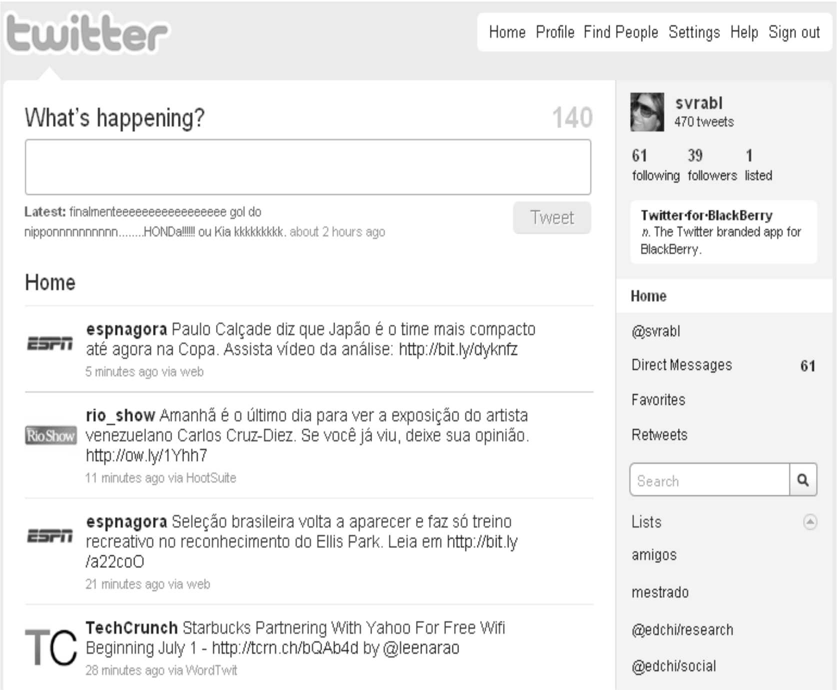
Basically, microblogging is a platform that allows sending messages with only 140 characters. It depends on users account logins in platform such as Twitter, our main reference once is the first and most popular (Watters, 2010).

Profile information (name, location, web address and short biography) can be inputted to offer more traces about user's personality and interests. Friend's selection is also required, when user starts following your friends and in future, can be followed by them. So, a friend is defined as a following or/and a follower, in a relationship that can be asymmetric. Thus any person that a user follows doesn't have to follow another user. Users can only manage a friend that is following in actions such as: group friends in a list, report to a spam, unfollow or block. This last option eliminates all the reciprocity and messages that has been sent. Then, the blocked person is eliminated from user's social network. This action has a stronger consequence than stop following the friend, which is also a follower once a user can stop following but friends can still receive this user's message. Other relevant aspect is that Twitter motivates the following engagement. If there is someone who starts following a user, and user doesn't want to, the only option is user blocks or report to spam.

Messages can be sent or received, and all are stored in public timeline (a sequence of messages of user, following and follower friends organized in a reverse time). All messages are public and the whole microblogging members can access them, they do not even need to be friends. When a message is received, a user can reply to another user or resend (retweet) to the followers, which will still appear in the user and the followers' public timeline.

Usually, a user replies friends' messages that were sent either or not by them. When a retweet occurs, it shows that the user believes that the message is interesting enough to be widespread to its followers. Reputation is important in social networking, so most often users are careful when resending information from others. All tweet messages cannot be reedited once sent to public timeline; the only option is delete after it. In order to analyze content, retweets must be eliminated, once it repeats the previous messages, but for connection analysis, it is relevant to be considered. The more people retweet user's messages the more it indicates that a user has important and relevant content to be seen. Private messages can be sent only to friends that are also its followers.

In Figure 1, the user's public timeline (messages with time and origin) is shown together with, the number of following and follower friends, the number of direct messages and tweets sent by user. In messages content, it shows the person or the company that sends the message and URL (starts with http). The symbol @ identifies people that are mentioned in the message.



Latest: finalmenteeeeeeeeeeeeeeeeee gol do nipponnnnnnnnnn........HONDa!!!!! ou Kia kkkkkkkkk. about 2 hours ago

espnagora Paulo Calçade diz que Japão é o time mais compacto até agora na Copa. Assista vídeo da análise: http://bit.ly/dyknfz
5 minutes ago via web

rio_show Amanhã é o último dia para ver a exposição do artista venezuelano Carlos Cruz-Diez. Se você já viu, deixe sua opinião. http://ow.ly/1Yhh7
11 minutes ago via HootSuite

espnagora Seleção brasileira volta a aparecer e faz só treino recreativo no reconhecimento do Ellis Park. Leia em http://bit.ly/a22coO
21 minutes ago via web

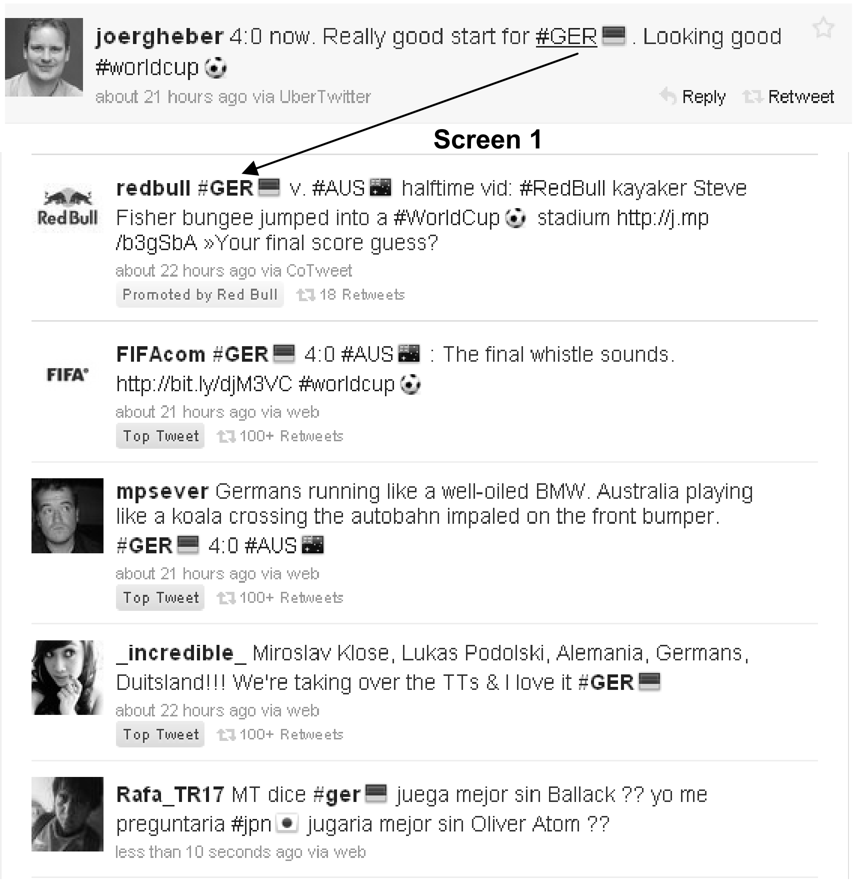
TechCrunch Starbucks Partnering With Yahoo For Free Wifi Beginning July 1 - http://tcrn.ch/bQAb4d by @leenarao
28 minutes ago via WordTwit

Twitter for BlackBerry
n. The Twitter branded app for BlackBerry.

Figure 1. User's Public Timeline of microblogging Twitter.

Messages can be set as favorite (by clicking in a star in the left side of message, which turns yellow) and can contain links (additional content represented by URL) and hash tags (words initiated with # that is an aggregation resource of interest topic and the following characters cannot have %! etc and no spaces. E.g.: #iloveworldcup). Figure 2 depicts an example of message that cites the hash tag #GER in screen 1.

When we click in the hash tag link, screen 2 shows all microblogging members (not necessarily user's friend) who posted messages with the same hash tag, in order to express opinions about the world cup game Germany x Australia. During that time, people communicate in various languages and much more through feelings and points of view about football game than discussing. The environment becomes a spontaneous place where people share the same interest in a specific given time. After the game is over, the hash tag #GER is less used and the interest moves to another #hash tag, possibly the next game, for example.



joergheber 4:0 now. Really good start for #GER . Looking good #worldcup
about 21 hours ago via UberTwitter

Screen 1

redbull #GER v. #AUS halftime vid: #RedBull kayaker Steve Fisher bungee jumped into a #WorldCup stadium http://j.mp/b3gSbA »Your final score guess?
about 22 hours ago via CoTweet
Promoted by Red Bull 18 Retweets

FIFAcom #GER 4:0 #AUS : The final whistle sounds. http://bit.ly/djM3VC #worldcup
about 21 hours ago via web
Top Tweet 100+ Retweets

mpsever Germans running like a well-oiled BMW. Australia playing like a koala crossing the autobahn impaled on the front bumper. #GER 4:0 #AUS
about 21 hours ago via web
Top Tweet 100+ Retweets

incredible Miroslav Klose, Lukas Podolski, Alemania, Germans, Duitsland!!! We're taking over the TTs & I love it #GER
about 22 hours ago via web
Top Tweet 100+ Retweets

Rafa_TR17 MT dice #ger juega mejor sin Ballack ?? yo me preguntaria #jpn jugaria mejor sin Oliver Atom ??
less than 10 seconds ago via web

Screen 2

Figure 2. Use of hash tags in Public Timeline of Twitter.

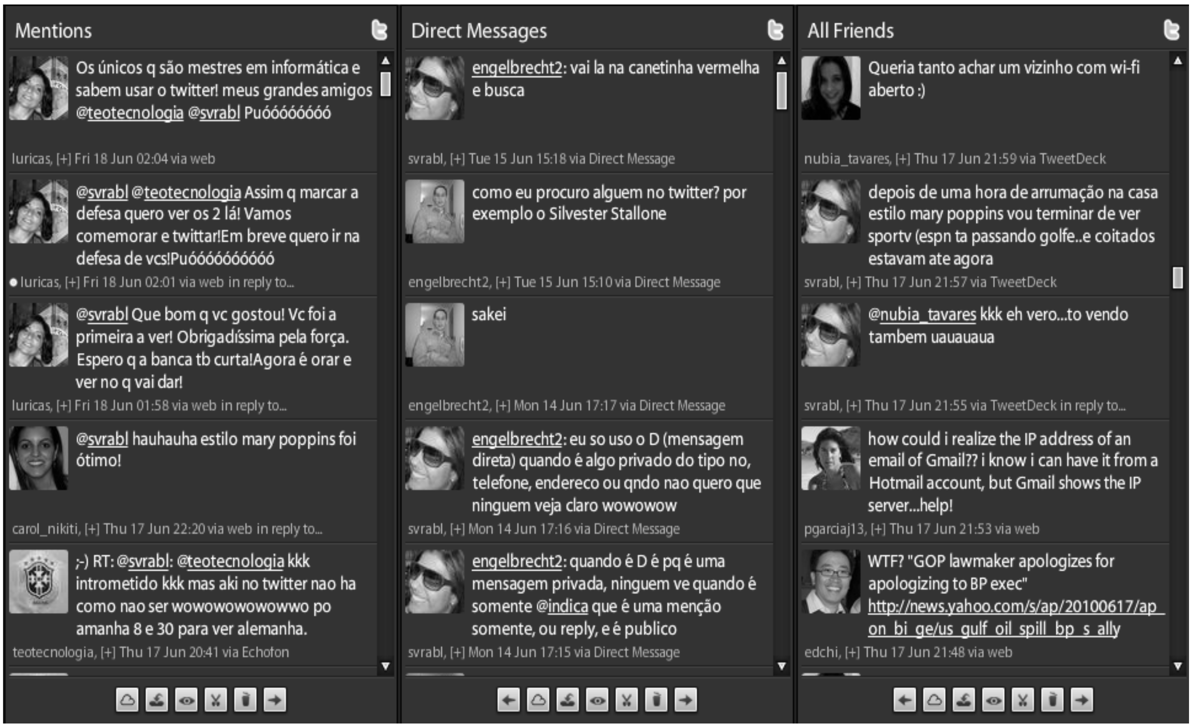
Once interests are volatile, tenuous, unstable, and temporary, encompassing different user's aggregations around them, it is important to understand much more the interest's relations beyond connections than the content exchange, once it can be easily obsolete. The hypothesis of this work concerns in exploring ways to optimize the process of acquiring information, once current resources don't consider the strength of social networking for monitoring interests: filters with manual search (e.g.: Google); alerts (e.g.: Google Alert); Registration by e-mail (Blog) or group (e.g.:Googlegroups); feeds (e.g.: RSS); social bookmarks (e.g.: Delicious), aggregators (e.g.: FriendFeed, Netvibes), metasearch (e.g.: Mr.Taggy), recommendation engines (e.g.: Amazon, Submarine).

2.1. Microblogging Dynamic Example

To better understand these dynamics, we performed a study observing the flow of Twitter messages during the event at the World Cup, the match between Germany and Ghana during the first stage of the championship.

We use the TweetDeck to do it so once Twitter interface doesn't provide visualization resources that combine multiple follows. And it seems not be Twitter's developers intention once they delivered an API to allow several third parties free services assist on facilities and personalization tasks. Tweet Deck is an example (Figure

3) and it is a desktop application where users can customize with columns, groups, saved searches and automatic updates helping user to effortlessly stay updated with the people and topics you care about. It can also seen what people are saying about you and join the conversation by tweeting, sharing photos, videos or links directly. Figure 3 shows columns that organize messages that mentions the user (left column), direct messages (center) and all public timeline (right), when user and user's friends messages are organized as the way as they appears.



Os únicos q são mestres em informática e sabem usar o twitter! meus grandes amigos @teotecnologia @svrabl Puóóóóóóóó
luricas, [+] Fri 18 Jun 02:04 via web

@svrabl @teotecnologia Assim q marcar a defesa quero ver os 2 lá! Vamos comemorar e twittar!Em breve quero ir na defesa de vcs!Puóóóóóóóóóó
● luricas, [+] Fri 18 Jun 02:01 via web in reply to...

@svrabl Que bom q vc gostou! Vc foi a primeira a ver! Obrigadíssima pela força. Espero q a banca tb curta!Agora é orar e ver no q vai dar!
luricas, [+] Fri 18 Jun 01:58 via web in reply to...

@svrabl hauhauha estilo mary poppins foi ótimo!
carol_nikiti, [+] Thu 17 Jun 22:20 via web in reply to...

;-) RT: @svrabl: @teotecnologia kkk intrometido kkk mas aki no twitter nao ha como nao ser wowowowowowwo po amanha 8 e 30 para ver alemanha.
teotecnologia, [+] Thu 17 Jun 20:41 via Echofon

engelbrecht2: vai la na canetinha vermelha e busca
svrabl, [+] Tue 15 Jun 15:18 via Direct Message

como eu procuro alguem no twitter? por exemplo o Silvester Stallone
engelbrecht2, [+] Tue 15 Jun 15:10 via Direct Message

sakei
engelbrecht2, [+] Mon 14 Jun 17:17 via Direct Message

engelbrecht2: eu so uso o D (mensagem direta) quando é algo privado do tipo no, telefone, endereco ou qndo nao quero que ninguem veja claro wowowow
svrabl, [+] Mon 14 Jun 17:16 via Direct Message

engelbrecht2: quando é D é pq é uma mensagem privada, ninguem ve quando é somente @indica que é uma menção somente, ou reply, e é publico
svrabl, [+] Mon 14 Jun 17:15 via Direct Message

Queria tanto achar um vizinho com wi-fi aberto :)
nubia_tavares, [+] Thu 17 Jun 21:59 via TweetDeck

depois de uma hora de arrumação na casa estilo mary poppins vou terminar de ver sportv (espn ta passando golfe..e coitados estavam ate agora
svrabl, [+] Thu 17 Jun 21:57 via TweetDeck

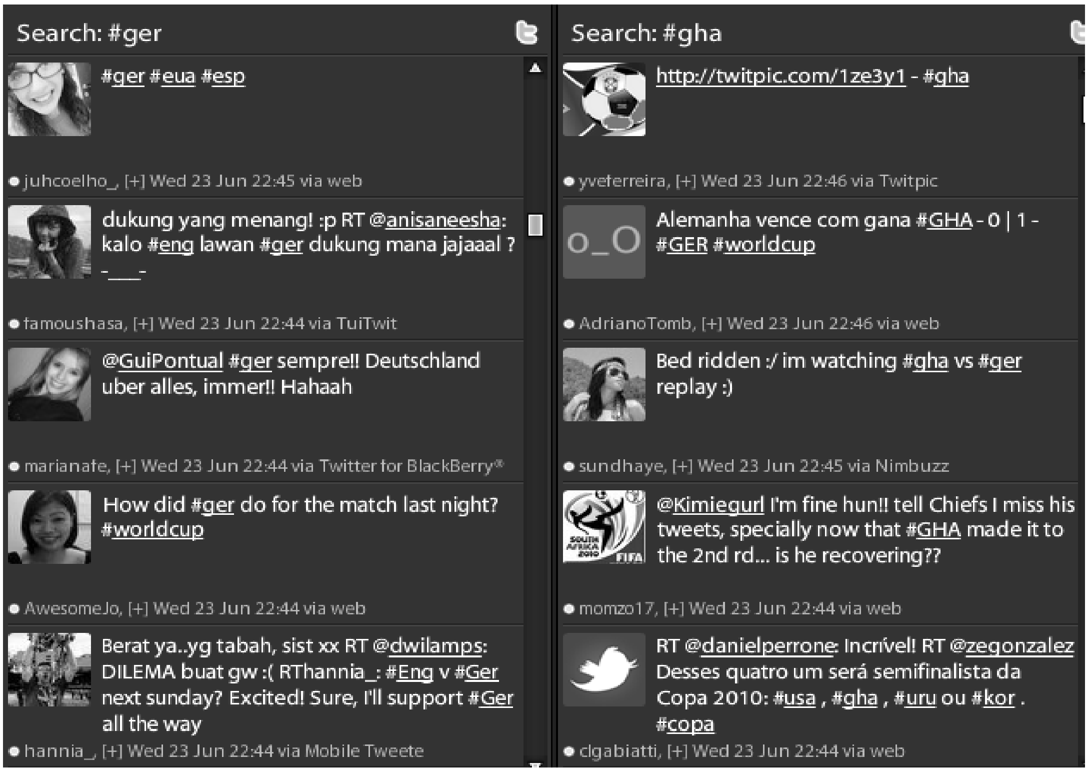
@nubia_tavares kkk eh vero...to vendo tambem uauauaua
svrabl, [+] Thu 17 Jun 21:55 via TweetDeck in reply to...

how could i realize the IP address of an email of Gmail?? i know i can have it from a Hotmail account, but Gmail shows the IP server...help!
pgarciaj13, [+] Thu 17 Jun 21:53 via web

WTF? "GOP lawmaker apologizes for apologizing to BP exec" http://news.yahoo.com/s/ap/20100617/ap_on_bi_ge/us_gulf_oil_spill_bp_s_ally
edchi, [+] Thu 17 Jun 21:48 via web

Figure 3. TwitterDeck Panel

In Figure 4, there are two columns were created in order to follow posts from: #ger, Germany team and #gha, Ghana football team. During event, the message flow is continuous and overwhelming. It can see that users are unknown, around a sharing interest: express themselves during the game and even after, with post of comments. Filter information is quite tough. For example: #ger #eua #esp in left first message only indicates that user wants to force that this countries be in top trending ranking tags list presented in Twitter´s main page. There are other messages that show only URL – the first message in right show that is a picture sent by @YveFerreira. The main challenge is identify messages patterns that are not considered in social match model during the interest filtering. Heterogeneity and data redundancy must be considered and also the understanding that users declare tags not just to classify its messages content, but express feelings, highlight them. For example, a message "I don´t like Football! #isuffertheconsequences." The tag is much more to stress some fun or feeling than classified the previous text. Other aspect is that in both columns there are people cheering for Germany and Ghana, so information relevant is much more in people than in content itself, which can be ambiguous. User can find an interesting message and retweet. The retweeted person will see it in mention´s column (Figure 3) and can trace if the person is interesting to follow or not. All process is manually analyzed, and we offer a perspective in helping filtering and identifying profile patterns around a specific interest.



Search: #ger

#ger #eua #esp

● juhcoelho_, [+] Wed 23 Jun 22:45 via web

dukung yang menang! :p RT @anisaneesha: kalo #eng lawan #ger dukung mana jajaaal ? -___-

● famoushasa, [+] Wed 23 Jun 22:44 via TuiTwit

@GuiPontual #ger sempre!! Deutschland uber alles, immer!! Hahaah

● marianafe, [+] Wed 23 Jun 22:44 via Twitter for BlackBerry®

How did #ger do for the match last night? #worldcup

● AwesomeJo, [+] Wed 23 Jun 22:44 via web

Berat ya..yg tabah, sist xx RT @dwilamps: DILEMA buat gw :(RThannia_: #Eng v #Ger next sunday? Excited! Sure, I'll support #Ger all the way

● hannia_, [+] Wed 23 Jun 22:44 via Mobile Tweete

Search: #gha

http://twitpic.com/1ze3y1 - #gha

● yveferreira, [+] Wed 23 Jun 22:46 via Twitpic

Alemanha vence com gana #GHA - 0 | 1 - #GER #worldcup

● AdrianoTomb, [+] Wed 23 Jun 22:46 via web

Bed ridden :/ im watching #gha vs #ger replay :)

● sundhaye, [+] Wed 23 Jun 22:45 via Nimbuzz

@Kimiegurl I'm fine hun!! tell Chiefs I miss his tweets, specially now that #GHA made it to the 2nd rd... is he recovering??

● momzo17, [+] Wed 23 Jun 22:44 via web

RT @danielperrone: Incrível! RT @zegonzalez Desses quatro um será semifinalista da Copa 2010: #usa , #gha , #uru ou #kor . #copa

● clgabiatti, [+] Wed 23 Jun 22:44 via web

Figure 4. Columns #GER and #GHA

3. Social Matching Model

As mentioned before, microblogging is a public environment where people express themselves in a free, spontaneous and fast way. Its potential in providing useful information depends on matching member's connections with member's interests. It means that if someone shares similar interests with me about one or more themes, it is high relevant to follow him/her, and after that, I have to evaluate if, in fact, this new "friend" has been providing insights about my interests. Whenever I judge that this relation is not relevant any more I can just stop following and stop receiving messages as well. But if I have a plenty of different interests inside and outside of established social network in Twitter, how this process can be optimized? Our proposal is implement a recommendation engine that will execute the following steps.

1.User assign to #twintera!

2.Engine will extract indicators not interest-related and will send to user

3.User will inform interest

4.Engine will look in user social network to identify who has more occurrences with the interest.

5.The user will pass through profile patterns identified by Bayesian net and will be selected (more details in Section 3.1)

6.The recommended result (Figure 5) will be sent by short URL and user will be click on a link and will see a map with users, profile and complementary interests. The main point is visualize the interest (In Figure 5, the example is Microblogging); people recommended (starts with @) and their complementary interests (starts with #). This complementary information is the key for user awareness of the reason of recommendation and other trendy interests that will be possible to share with. It helps also user analysis of people recommended without accessing their profile and public timeline.

7.Check if next 24 hours user follows some person recommendation and 48 hours if unfollow. It avoids that people use strategies to create artificially more followers (once it means more status) such as: user follows a person, then the person follows back. One day later, user unfollows that person, who keeps the follower relation.

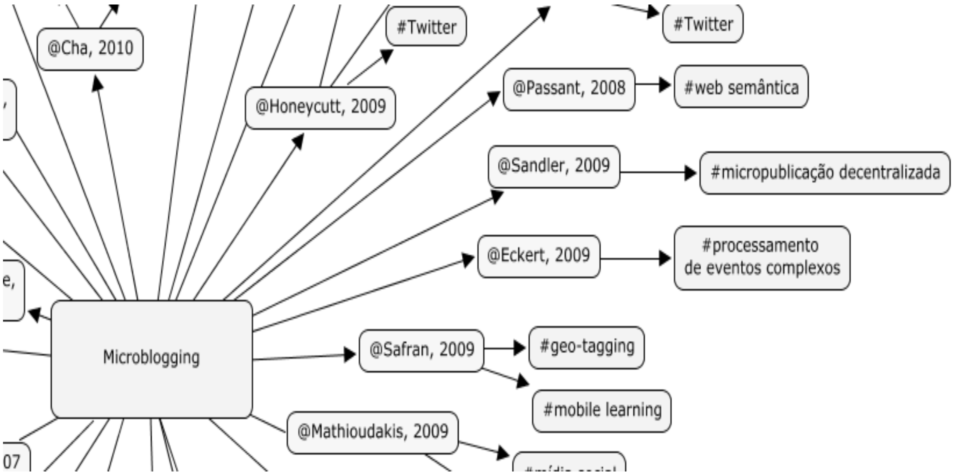


Figure 5. Recommendation Map Result

3.1. Knowledge Indicators

In order to provide more accurate recommendation we need to identify the level of knowledge of user. It is a way to identify his state in microblogging and its relevant contribution. We propose three levels of knowledge profiles: enthusiasts, specialists and apprentices. To set them, we mapped indicators from user's social networks activities and functionalities in Twitter.

In Table 1, there are the indicator´s description and an example of three profiles from @svrabl public timeline in Twitter: @svrabl – enthusiasts, @engelbrecht2 – apprentice and @edchi – specialist. Depending on the interest, the user can be a specialist, enthusiast or apprentice. All profiles were selected freely. As the data is constantly changing, we are taking stats for day 18th June 2010. The example numbers for each profile x indicator were taken for third party services that uses Twitter API and from Twitter itself. During the indicators example process we realize that there is no a social matching engine that encompasses our entire indicators proposal. We believe that is due to Twitter offers some limitations in traffic for using your data, issue that we will have to deal in implementation phases.

Indicators are spread in many services (see the footnotes). The values TBE (to be extracted) means that we need to implement a program to identify those numbers in order to analyze, select data and establish patterns.

Our hypothesis is that specialists are partially active, with homogeneous interests, shared in a much reduced number of messages which explore more professional activities. Enthusiasts are very active, with heterogeneous interests, shared in very large number of messages, explored emotions which also describes more activities and feelings. Usually using hash tags with personal meaning (# iconfessthat). Apprentice is a newcomer in some interest that starts your relationship with friends in a smaller group.

Ouslavirta et al (2009) identified in Jaiku microblogging that enthusiasts are *more active*, (average of 212 posts sent), *better networked (*around of 35 following and 38 followers) and *veteran user (average of* 85 days of retention). T*heir practices are extremely interactive*: "Although they make up only 2.5% of the user population, they are responsible for 30% of all Jaiku (messages) and 46% of all comments sent. In addition, they likely influence and receive attention from the non-enthusiasts."

Table 1. Indicator Data x Profile Traces

Aspects	Indicators[1]	Description	@svrabl – enthusiasts	@engelbrecht2 – apprentice	@edchi - specialist
Profile	**Link**	Additional content field "Web" in Twitter	http://svrabl.wordpress.com/	No specified	http://www.edchi.net
	Biography	Field with 160 characters containing user self-description or explicated interest.	# Msc Thesis #microblogging, #social education, #informal learning #social matching #recommendati	No specified	Augmented Social Cognition; Area Manager at PARC; HCI and Social Computing Researcher

[1] Information extracted manually from http://twitter. Those exceptions are informed.

Aspects	Indicators[1]	Description	@svrabl – enthusiasts	@engelbrecht2 – apprentice	@edchi - specialist
			on systems #bayern muenchen #mario gomez #flamenco #tchibo		
	Location	User´s physical geography	Rio de Janeiro, Brazil	No specified	Palo Alto, California
	Joined on[2]	Date of registering on Twitter	2009-02-04	2010-06-03	2007-10-05
	Grade[3]	Impact analysis. Evaluates power, reach and authority of a twitter account	65%	29%	96%
	Following	Number of Twitter others users that @user subscribe to follow their Tweets or updates on the site	63	2	286
	Followers	Quantity of other Twitter users you have chosen to follow on the site	41	3	686
	Listed	Quantity of another Twitter user's list that @user is included	2	0	80
	Tweets	Quantity of a message posted by @user in Twitter	616	18	827
Trustiness	**Spam**	Do not consider people that have only symbols in profile. Also discard people with many followers and followings, but with few messages. Do not consider sex links.	See Figure 6	See Figure 6	See Figure 6
Reciprocity[4]	**Following**	Quantity of @user's following these people, but they're not following @user back.	36	1	112
	Fans	Quantity of people	15	2	512

[2] Using results from http://twitter.grader.com/

[3] Using results from http://twitter.grader.com/

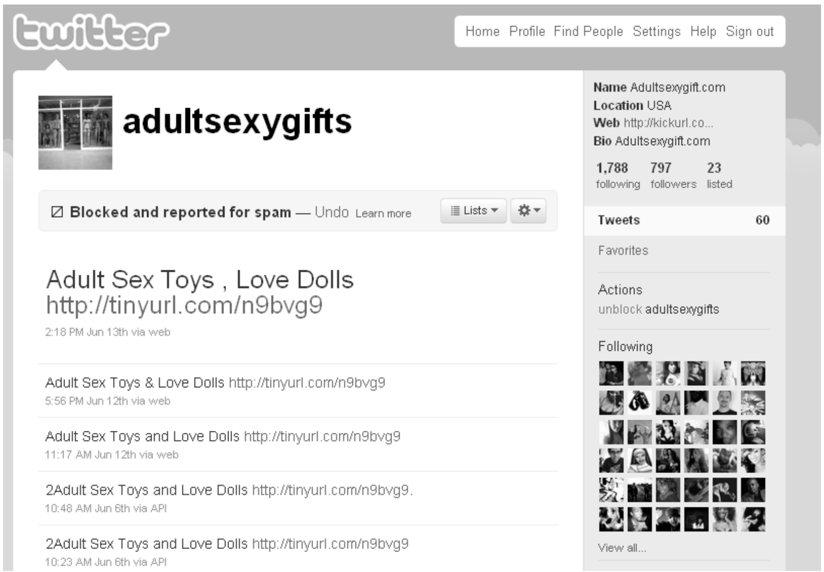
[4] Using results from http://friendorfollow.com/

Aspects	Indicators[1]	Description	@svrabl – enthusiasts	@engelbrecht2 – apprentice	@edchi - specialist
		who are following @user, but @user not following them back.			
	Friends	Quantity of people who are following @user and @user's following them back.	24	1	174
User Activity	**Inactive[5]**	People you are following who hasn't updated their Twitter status (tweeted) in the past 30 days	18	22	0
	Active	% active people that user follows and is also a follower	TBE	TBE	TBE
	Followers	Quantity of follower messages sent during a day	TBE	TBE	TBE
	Following	Quantity of following messages sent during a day	TBE	TBE	TBE
Tweets Timeline[6]	**Tweets Day**	Average of Nr. Tweets per day	7	2,5	3,7
	Tweets Month	Average of Nr. Tweets per month	53	18	54
	Retweet	% of total tweets @user retweet	1,52% of total tweets	No specified	31,89% of total tweets
	Replies to	% of total tweets @user replies to	25,8% of total tweets	22,2%	27,63% of total tweets
	Total Messages	% message sent x receipt during a day	TBE	TBE	TBE
New Associations	**Followers**	Quantity of new followers	TBE	TBE	TBE
	Unfollowers	Quantity of new unfollowers	TBE	TBE	TBE
	Active Followers	Quantity of new followers and that is also active	TBE	TBE	TBE
	Active Unfollowers	Quantity of new unfollowers and that is	TBE	TBE	TBE

[5] Using http://twitoria.com

[6] Using http://tweetstats.com/

Aspects	Indicators[1]	Description	@svrabl – enthusiasts	@engelbrecht2 – apprentice	@edchi - specialist
		also active			
Popularity	**Mentions to user**	Quantity of messages that mentions @username	TBE	TBE	TBE
	People that retweets most to user	Identify who retweet more to user. Check if they are in user's friends.	TBE	TBE	TBE
	People that user most retweet	Identify who are those user retweet most. Check if they are in user's friends.	TBE	TBE	TBE
	Favorite friends	Identify who are those ones that user indicates a favorite message	TBE	TBE	TBE
Interests By User	**Volume of # from @user**	Identify at the same time, the hash tags and people more cited in user messages	TBE	TBE	TBE
Interests By Popularity	**Volume #**	Quantity Hash tags that user most uses	TBE	TBE	TBE
Interests By Frequency	**New #**	New Hash tags that user most uses	TBE	TBE	TBE
Interests By Additional content	**Volume de links**	Quantity of links sent by user, not considering retweet messages.	TBE	TBE	TBE
Interests by Repetition Grade	**Retweets/Total of Messages**	%of messages that user retweets/total of messages sent by user.	TBE	TBE	TBE
Interests by # activity	**# hashtags more used in a given moment**	Compare if the user #hash tag in sent messages is in also Twitter main tags	TBE	TBE	TBE
Interests by inactivity grade	**#hashtags in time**	Time that a hash tag was previous used, than turns inactive, and then returns to message sent by user.	TBE	TBE	TBE



adultsexygifts

☒ Blocked and reported for spam — Undo Learn more

Adult Sex Toys , Love Dolls http://tinyurl.com/n9bvg9
2:18 PM Jun 13th via web

Adult Sex Toys & Love Dolls http://tinyurl.com/n9bvg9
5:56 PM Jun 12th via web

Adult Sex Toys and Love Dolls http://tinyurl.com/n9bvg9
11:17 AM Jun 12th via web

2Adult Sex Toys and Love Dolls http://tinyurl.com/n9bvg9.
10:48 AM Jun 6th via API

2Adult Sex Toys and Love Dolls http://tinyurl.com/n9bvg9
10:23 AM Jun 6th via API

Name Adultsexygift.com
Location USA
Web http://kickurl.co...
Bio Adultsexygift.com

1,788 following 797 followers 23 listed

Tweets 60

unblock adultsexygifts

Figure 6. Example of Spam Profile in Twitter

6. Related Works

Degirmencioglu et al (2010) proposes a more efficient model to identify relevant content in microblogging, comparing what people report and their actual contributions, based on the premise that the most popular are not always the highest contributors. To do so, it ignores the explicit categorization of the user (lists, groups, etc.), processes information, reducing them to keywords that represent the nature of the content. Then, identifies the community of interest based on the user.

There are third party services to facilitate access to information of microblogging. For this study, we identified three main areas, shown in Figure 7.

We stress as its main limitation the fact that most recommendations of people is obtained by crossing common friends of the network where the user does not inform the subject of interest. The results are not always explained, but when they are, come in a list format and users have to manually identify the complementary interests of each. Furthermore, the search for information requires the user to clarify your topic of interest, and we consider this assumption in our model.

There are also services that use forms of maps, graphs and arcs to show the timeline of the messages. There is no, however, a unique service that combines all of these perspectives and, therefore, we propose the interaction #! as a prototype to serve this purpose.

Silva (2009) presents a model of social combination that appoints people based on their similar interests in a social bookmarking tool. Differs from the work proposed here by the characteristics of the environment, the method of viewing the results of the recommendation be on lists, not offering additional interests from persons recommended by and adopt a heuristic method as Pearson's correlation.

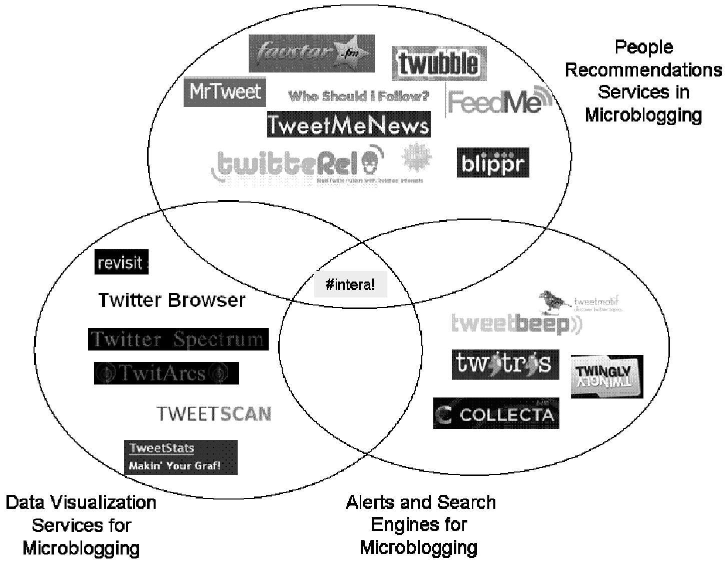


Figure 7. Microblogging Services and #twintera!

6. Conclusion and Future Works

Microblogging is not just a trend, but an ecosystem platform that integrates people and provides constant and fluid information. Its adoption has increased publications significantly on this subject. We presented how to use and what are the main possible actions in the platform, using Twitter as reference.

We also describe our social matching model, as a way to provide information search by connections, rather than focusing in content. We propose indicators to identify profile and levels of knowledge, aiming to provide people premium recommendation.

The model will recommend people with similar interests and show their complementary ones and it is planned to be implemented using Twitter API.

The future works firstly consists on refining profile and level of knowledge indicators, identifying patterns. Secondly, the recommendation engine will be validated in an experiment in order to confirm that it can provide enhance knowledge acquisition by amplifying connections and approaching people with similar interests.

References

Barkhuus, L.; Brown, B.; Bell, M., (2008). From awareness to repartee. In: Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08. Anais. p.497. Florence, Italy. doi: 10.1145/1357054.1357134.

Degirmencioglu, A. E. ; Uskudarli, S. (2010) Exploring Area-Specific Microblogger Social Networks. In: Proceedings of the WebSci10: Extending the Frontiers of Society On-Line, April 26-27th, 2010, Raleigh, NC: US.

Java, A.; Song, X.; Finin, T.; Tseng, B., (2007) Why we twitter. In: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis - WebKDD/SNA-KDD '07. Anais. p.56-65. San Jose, California. doi: 10.1145/1348549.1348556.

Krishnamurthy, B.; Gill, P.; Arlitt, M., (2008) A few chirps about twitter. In: Proceedings of the first workshop on Online social networks - WOSP '08. Anais... . p.19. Seattle, WA, USA. doi: 10.1145/1397735.1397741.

Miller, V., (2008). New Media, Networking and Phatic Culture. Convergence: The International Journal of Research into New Media Technologies, v. 14, n. 4, p. 387-400. doi: 10.1177/1354856508094659.

Oulasvirta, A.; Lehtonen, E.; Kurvinen, E.; Raento, M, (2009). Making the ordinary visible in microblogs. Personal and Ubiquitous Computing. doi: 10.1007/s00779-009-0259-y.

Passant, A.; Hastrup, T.; Uldis, B.; Breslin, J., (2008). Microblogging: A Semantic and Distributed Approach. Proceedings of the 4th Workshop on Scripting for the Semantic Web (SFSW '08). Tenerife, Spain.

Silva, S.P de A., (2009) Oraculous: Um Modelo para Combinação Social em Redes Sociais. Dissertação (Mestrado em Informática) – Universidade Federal do Rio de Janeiro, Instituto de Matemática, Núcleo de Computação Eletrônica, Programa de Pós-graduação em Informática.

Sutton, J.; Palen; Leysia; Shlovski, I., (2008). Back-Channels on the Front Lines: Emerging Use of Social Media in the 2007 Southern California Wildfires. Proceedings of the 2008 ISCRAM Conference. Washington, D.C.

Voida, A.; Voida, S.; Greenberg, S.; He, H. A., (2008). Asymmetry in media spaces. In: Proceedings of the ACM 2008 conference on Computer supported cooperative work-CSCW '08. Anais. p.313. San Diego, CA, USA. doi: 10.1145/1460563.1460615.

Watters, A., (2010). Just the Facts: Statistics from Twitter Chirp. Available in:<http://www.readwriteweb.com/archives/just_the_facts_statistics_from_twitter_chirp.php>. Access in 16th April, 2010.