“All your data are belong to us” or how the Internet giants may have a card up their sleeve.

Luis MartinGarcia, Joaquin Salvachua.

Departamento de Ingeniería de Sistemas Telemáticos.
Escuela Técnica Superior de Ingenieros de Telecomunicación.
Universidad Politécnica de Madrid.
{luis.mgarcia@alumnos.upm.es, jsalvachua@dit.upm.es}

***Abstract* — While many people are concerned about the amount of data Internet giants like Google or Facebook have acquired over the years, it is still not clear how much of that data are they using or what they are using it for. Also, it seems to be a common belief that companies of the digital age are only present in the digital world, and that only traditional companies may operate on both sides. This paper will discuss the kinds of data Internet giants have access to, and will present some examples on how to use that data for profit, showing how Internet-based companies may jump into traditional business markets, just like traditional companies have moved part of their business to the on-line world.**

I. INTRODUCTION

Very few people would disagree that the Internet has changed the world in a way comparable to the introduction of the printing press or the alternating current electricity. Since its origins, the Internet has provided humans with new ways to communicate, and has been the most important piece of the so-called Information Age. It is true that the net is currently not equally accessible from all parts of the world, but as far as developed countries are concerned, a very significant percentage of the population performs on-line transactions on a daily basis. Among other things, this has changed the way humans access knowledge and the way they communicate and interact with each other. Sources of information like books or personally known people are being progressively abandoned in favour of digital contents accessible through a web browser. A number of people gatherings and phone conversations also seem to be getting replaced by on-line chats and social networking sites. Human behaviour is changing rapidly and it seems that the cultural gap between generations is increasing faster than ever.

The Internet, however, may not be the only responsible for these changes, as the behaviour of its users has varied significantly over the years. To take an example, while in the 90s, Internet users tried to avoid exposing personal information to the net, hiding themselves behind pseudonyms (handles or nicknames in the Internet jargon), and making up false identities; in the last decade users have become just the opposite, and are now willing to disclose every aspect of their lives to their favourite social network or the readers of their blogs. This, as I will discuss later, has some implications.

Apart from people's personal lives, the digital revolution has also had an enormous impact on businesses. Companies have always been successful in adopting inventions and technologies that allowed cutting costs, and the Internet is no exception. The on-line world has provided companies a cost-effective way to reach more customers, and has provided local businesses an easy way to expand their potential market world-wide. Examples include the financial, information and service sectors. In the same way, a significant amount of retail companies have moved part of their distribution channels to the Internet [1] and also, manufacturers have started to complement their traditional retail paths with direct on-line channels [2]. Note that this trend is not reflected equally in all sectors, but it seems clear that, on average, the use of the Internet to expand or maintain business models is constantly increasing. However, Internet-born companies do not seem to follow the analogous path extending their presence off-line. Internet giants like Google and Facebook limit their market to the on-line world, even though they could obtain substantial gain from the other side. In this paper, we analyse how these companies could extend their activities to traditional markets, using their user base and the information obtained from it as a key asset to increase their profits.

The remainder of this paper is organised as follows. Section II describes the two companies that are subject of this study: Google and Facebook. Section III describes the information that those companies may have access to as a result of their significant user base. Section IV briefly discusses which pieces of data may be unknown to them and how some of it could be inferred from their existing data bases. Section V discusses the possibility of using the data for profit, presenting some simple real-world examples on how they could turn their knowledge into money by either selling it or using it to invest on particular assets. Finally, section VI presents the author's conclusions, and discusses open questions and future work.

II. THE INTERNET GIANTS

The Internet is a huge but fast-changing market where many companies have their 15 minutes of success and then disappear. However, there are a few companies that seem to be there to stay: the Internet giants. In this paper we will focus on the two most important ones: Google and Facebook, which are the first and second most accessed web sites respectively [3].

Nevertheless, most of the ideas that will be presented could also be extrapolated to other companies with a big presence on the Internet, such as Microsoft, Twitter or Linkedin.

Since its foundation in 1998, Google has become the de facto web search engine on the Internet. There are no official statistics on the use of the Google search engine, but some sources estimate that it receives more than 10 billion search queries per month only in the US, which represents a local market share of 65% [4]. It's on-line video streaming service, YouTube, serves around two billion videos a day worldwide and its users upload approximately 24 hours of video every minute [5].

Facebook is a younger company but its user base has grown rapidly since its website was launched in February 2004. As of January 2011, Facebook claims to have more than 500 million active users (users who have returned to the site in the last 30 days) [6], although other sources suggest that the number may be closer to 600 [7].

Both companies receive most of their revenue from their advertising programs: Google has the Adwords and Adsense programs, which account for 97% of Google's revenue [8] and Facebook has Facebook Ads, which also represents its primary source of income.

III. WHAT THEY KNOW

A. Google

A.1 Web data

Google indexes billions of publicly accessible web pages. In practical terms, this means that it has local direct access to an enormous pool of data. It is true that these data can theoretically be gathered by any other company or individual, because it can be accessed through the WWW, but Google has a competitive advantage over others, as it has a very powerful infrastructure that provides a huge amount of storage and the ability to continuously gather new data or update existing information. To take an example, back in 2006 the size of their web crawling database was approximately 850TB [9].

A.2 Query logs

Apart from the data extracted from the web itself, Google has access to a more powerful set of data: the information collected from its users through their use of the different Google services. Perhaps the most basic piece of information that a search engine has access to is the record of the terms their users searched for. These are commonly known as query logs and can be classified in two groups:

- Searches for textual information: These are the most common searches. They usually come from Google's main search engine (the popular google.com interface), but may also be obtained from those sub-engines that restrict searches to a subset of the total index. Examples include Google Scholar, for academic papers; Google Code Search, for software source code; Google Finance, for stocks, mutual funds, and public and private companies; or Google News, for newspapers and other news sites.

- Searches for non-textual information: Apart from the traditional search engine, Google provides a few services that allow users to conduct searches for multimedia contents. To name a few examples, Google Video and YouTube for music clips, movie trailers and amateur videos; Google Maps for geographic locations and services offered near those locations; or Google Images and Google Picasa Web Albums for photos and other types of images.

A.3 Search result clicks

When Google's search engine receives a query, it responds to the user with a list of links to web sites, or other resources like maps or videos, that are relevant to such query. Users typically skim through the search results and click on one or more of the links. The traditional HTTP/HTML technology does not allow website owners to determine which links their users clicked on. However, there are certain tricks that can be used to obtain that information. In particular, Google does not provide direct links to the sites on the results, but links to an internal URL redirection system that forwards users to the web site they expected while allows the search engine to keep a record of the results that users clicked on.

A.4 E-mail and chat logs

The Gmail service, also known as Google Mail, has about 200 million users [10]. By default, all messages that are received or sent from a Gmail account are kept in Google's servers. This allows Google to access the data in their user's email accounts. While legally it is not possible to associate message data with a particular user, Google scans email messages for certain keywords in order to add context-sensitive advertisements to them. If statistics for the ads displayed are kept, either directly or indirectly, Google would have aggregate statistics for the words that are appear in all e-mail messages.

Google also offers an instant messaging service called Gtalk. It allows basic live textual communication as well as voice and video conferences. Logs for text chats are stored unless the user opts out explicitly. There is no evidence that Google stores voice and video communications.

A.5 Geographical location

Any website owner can determine a visitor's approximate geographical location based on its IP address. There are freely accessible geolocation databases available that allow to associate IP address with locations at a city level [11]. In addition, Google has a geolocation service, called Google Latitude [12], that lets users share their exact location with others at all times. This is achieved by having an application that runs continuously on the user's access device, which continuously updates the location of the user. The location can be determined from various sources. Examples include, the IP address, GSM access cell, Wi-Fi-based positioning systems, or GPS coordinates, many of which are provided natively by modern cell phones and other mobile devices.

B. Facebook

B.1 Personal details

In order to register for a Facebook account, users need to supply their first and last name, e-mail address, gender and date of birth. Additionally, on the first time a user logs in, Facebook provides a step-by-step wizard to configure the account which, among other things, encourages users to provide other personal information like the high school and university they attended or their current employer.

Facebook also lets its users extend their profile information by providing complete forms to supply all kinds of details. Users may tell Facebook about their city of residence, their hometown, their sexual orientation, the languages they speak, company projects they have participated in, their religion, their political beliefs, people that inspires them, their favourite quotations, music, books, movies, television shows or video-games, their activities and interests, alternative e-mail addresses, instant messaging screen names, phone numbers, full address, personal website, etc.

B.2 Connections

One of the most important characteristics of Facebook and other social networks is that they provide mechanisms in which users can establish connections with each other. In Facebook, these connections are called friendships, and require both ends to agree in order to be established (one end requests the friendship and the other accepts it). Apart from friendships, Facebook also lets its users specify other kinds of relations with others. Examples include love relationships, marriages, engagements, family relatives, etc.

B.3 Status updates and "liked" elements

Part of Facebook user's daily life consists of updating their status by answering the question "What are you doing right now?" through a simple text box displayed on their profile's front page. Also, it is typical that users express their interest on all kinds of elements (comments, people, groups, pages, etc), by clicking the "Like" button. These interests are immediately shown in their profile information, where their connections can see them.

B.4 Wall messages, private messages and chat logs

Facebook provides at least three channels in which users can communicate with each other. One of them is "the wall", which is a space on each user's profile page that allows friends to post messages for the user. Those messages are visible to anyone with permissions to access the user's full profile. Users may also exchange private messages that are only accessible by the sender and the receiver. Finally, users may perform live one-to-one communication through the Facebook Chat. The information exchanged from the two first channels is permanently stored in Facebook central servers, but there is no information available on whether Facebook stores the chat logs or discards them after conversations have ended.

B.5 Geographical location

Facebook has a service similar to the Google Latitude service discussed above. Its name is Facebook Places [13] and lets users provide their current geographical location (typically using a mobile device), so their friends know where they are at that moment. The service also allows users to report the presence of other users on the same spot.

IV. WHAT THEY DO NOT KNOW

From the discussion presented above, it seems clear that, due to the size of their user base, companies like Google or Facebook possess an important amount of information that very few other companies could have access to. However, that information is not entirely complete.

It could be said that Google knows the information that is available on the web, and also, and more importantly, the interest of people in that information (determined through query logs and search result click logs). Although there is a significant amount of information that can be inferred from such data [14, 15], it is not clear whether Google is able to build accurate

profiles of its users. On one hand, many of their services do not require users to be logged in to use them. Google makes efforts to compensate this, by setting cookies that allow user tracking between sessions, but it certainly does not guarantee accurate information. Even when users are logged in, the information that these users provide at the time of registration for a Google Account merely consists on their email address, a password and their country or residence. Therefore, in order to build more complete profiles of its users, Google needs to infer information from the data that they generate. One example would be inferring the user's gender from his voice. Google "search by voice" is becoming very popular in mobile devices where keyboard typing is slow. Also, Gtalk offers voice services. By analysing the user's voice, it may be possible to tell men and women apart, separate kids from adults, or even determine the user's geographical region based on his spoken language [16, 17, 18, 19]. Gtalk video conferences may also be a good source of information for age and gender determination.

There is also the case where a single computer is shared by more than one user. If all users access services using the same cookies or the same account, the profile, as seen by Google, would be inaccurate, as it would reflect the aggregate information of whole group of users. However, there is one case where it may be possible to differentiate those users: Google's search engine with Google Instant enabled. Google Instant is a feature that allows the search engine to start displaying results from the moment the user types the first letter of the query. As the user types, his keystrokes are sent to Google, where they are processed in order to generate the list of results for the partial query. Some authors suggest that humans may have unique (or at least unique enough for some purposes) keystroke patterns [20]. If this proves to be true, Google could analyse the way users enter their search queries in order to determine who is behind a computer keyboard or even a smart-phone pad [21].

Facebook, on the other hand, does not have accurate information on what its users are interested in, because they do not "ask" Facebook for information, just like they do with Google. One potential source for this kind of data could be the status updates. Some users may update their status with information about what they are interested in at that moment, products they wish to buy or have bought, or to express their opinion on a specific matter. Wall messages could also be processed to extract similar information. However, in order to be useful, that kind of data would need to be filtered, as users are also likely to update their status with vague sentences about their current mood or their daily routines. "Liked" items could be another source of information, but it would also require some filtering, as anecdotal evidence suggests that apart from real brands and products, users tend to like pages or messages containing jokes, challenges or some other useless information.

As opposed to Google, Facebook does actually have a fairly complete profile of its users, and it is in a better position to identify groups of people. Pieces of information like gender, age, place of birth or studies that are hard to obtain for Google, are easily accessible by Facebook, because they are supplied by its users in the registration process. Even when some pieces of information are missing, it is possible to infer a number of a user's attributes from the attributes of his connections [22].

Obviously Facebook cannot trust everything its users say about themselves but it can probably discard false information easily. To take an example, some authors suggest that it is possible to detect users that report false ages, based on the ages of their social connections [23].

V. USING KNOWLEDGE FOR PROFIT

After discussing which pieces of information the two Internet giants may have access to, we present a few ideas on how those companies can use the data for profit in the off-line world. For this, we make two fundamental assumptions:

- User data is not used individually, but in an aggregate form. In other words, the data is not associated with any concrete individual, but reflects groups of anonymous users.

- Google and Facebook make profit from their data either by becoming a consulting firm that sells the information to a third party or by using the data as an indicator to make certain investments. Note that Google already offers access to certain pieces of such data through the Google Insights for Search service [24]. In order to adopt some of the proposed strategies, Google would need to shut down that service.

A. Trend spotting

Market trends are the upward or downward movement of a given market, during a period of time. The ability to discover market trends and consumer patterns has been of interest in the area of market research for some time. There are many companies that sell detailed reports on the current state of certain markets. Google and Facebook could easily compete with them, using their internal information to produce similar (and possibly more accurate) reports. To illustrate this approach, we present two cases.

- The first one is the case of a breakfast cereal manufacturer. The marketing policies carried out by these companies, whose main target market are kids, focus on the creation of attractive packages for that kind of consumers. For this reason, these companies may be interested in knowing who the most popular cartoon character of the moment is, so they can contact its rights owner and establish an agreement to include the cartoon in the packaging of their products.

- In the same way, the audience share of a late night show is influenced by the popularity of their guest stars. As a consequence, producers may be interested in comparing the popularity of various people, in order to choose the best guest for the show in terms of share.

User generated information may not only offer information on market trends, but also on the personal habits of people. Of particular interest may be the information gathered by the Google Maps service. Query logs for this service, along with other geolocation data, provide important information about the mobility habits of its users. An analysis of this data may reveal things like transportation mean preferences, popular routes within a city or a country, abnormal behaviours, etc [25]. This knowledge may be useful at different levels. For example, popular routes at the city level may be useful to perform or complement traffic analysis and public transport route optimisation. Additionally, if Google could match locations and routes with other data like the gender or the age of the users that requested the maps, such information could be used by businesses seeking to open new shops in a city to select the better spot. At higher levels, the information could be used by airlines or tour operators to adapt their destination catalogue to the current demand.

YouTube's search query and result click logs may be of interest for the content, marketing and advertising industries. Factors like the number of views, the virality of a video, or the visualization patterns of users, can be analysed automatically [26] and may provide useful information for the creation of audiovisual content or advertising campaigns.

B. Target market determination

The term target market can be defined as the group of customers that a business aims its marketing efforts. Determining the optimal target market for a product is of great importance for many businesses, specially for small and medium-scale companies that cannot afford big marketing campaigns.

Google search query logs may be used to determine which geographical regions have more interest in a given product (or kind of product). Companies seeking to expand their presence in other regions could be interested in this service. For example, a company selling tea-based drinks may analyse query logs for terms like “tea”, “coke” or “juice”, for hints on the popularity of the different types of drinks in each region, and decide to set up their new base in India, rather than in Mexico. In the same way, a British tour operator specialised in extreme sports holiday packages may find that Australia and New Zealand are the best targets for their next marketing campaign, based on the interest on keywords like “bungee jumping”, “rafting” or “skydiving”.

Facebook “likes”, status updates, and profile data, may be useful to discover the consumer's gender and age ranges for a given product. Companies that carry broad scope marketing strategies may chose to reduce their scope, based on that information. For example, a gaming company may find that their video-games are consumed by males with ages between 10 and 18 and ages between 26 and 35, and therefore, decide not to target college students. This could also reveal unexpected target markets. A company targeting males between 18 and 25 may discover that their product is also being used by women in their 50s, and therefore, extend their marketing efforts to cover such segment.

C. Stock prediction

Stock price variations are a direct result of the supply and demand for those stocks. Google is in a good position to determine the interest of many people around the world on various assets. For example, Google could use their query logs (from either their main search engine or the Google Finance service) to measure the interest of its users in certain stocks, bonds or real estate properties, and compare that data to the real price of the assets. This could reveal leading indicators of future price variations that could be used to purchase options for the acquisition of relevant assets.

Google is aware of the potential use of their query logs to build forecasting models [27, 28]. Moreover, there is informal evidence that it has considered the possibility of performing stock market prediction, but decided to discard the idea as it could be illegal [29].

D. Hype cycle and product life cycle analysis

Products progress through a sequence of stages, from introduction to growth, maturity, and decline. This sequence is known as the product life cycle (PLC) and is associated with changes in the marketing situation, thus impacting the marketing strategy and the marketing mix. Additionally, new technologies and releases of innovative products typically involve certain levels of over-enthusiasm or "hype" and a subsequent disappointment of the customers, followed by a period of enlightenment and stability. This process is known as the hype cycle (HC) [30].

The development of forecasting methodologies to accurately predict PLCs and HCs could be of help in strategic decision-making for high-technology firms, research institutes, governments, universities, etc. The Internet giants could try to develop such methodologies, analysing data like search query logs or status updates, and attempting to find a correlation between such data and the actual life and hype cycles of a product. If such correlation exists, it may be of interest to the product manufacturer, its competitors, or even investors seeking to acquire distribution rights of a product in its growth stage.

In the same way, data mining and opinion gathering techniques could be applied to Google's web index data in order to extract information on aspects like consumer satisfaction, price perceptions, etc.

E. Corporate and industrial espionage

In the same way that individuals progressively abandon traditional communication channels like the e-mail or the telephone in favour of social networking sites like Facebook or Twitter, it seems reasonable to assume that this trend will also be reflected in the way business perform their B2B communications or in the way their employees communicate with each other [31]. This has severe security implications as the control of the data is shifted from the companies to a third uncontrolled party: the social networking sites. This situation may be exploited by these sites to gain inside information about a company's trade secrets, future plans, contacts, etc. Of course, this activity may be illegal, but there is certainly a strong temptation.

Google may be able to access similar information through Google Apps for Business, which provides services like e-mail, calendar and on-line document storage to private companies. However, there is another interesting source of information: query logs. Big companies typically have fixed IP address ranges assigned to them, and the assignments are normally public knowledge. For this reason, Google could theoretically group search queries by the company they were issued from, what would probably reveal interesting information on what those companies are up to.

VI. CONCLUSIONS

In this paper we have described part of the information that companies like Google or Facebook have access to, as a result of their enormous user base. We have shown how Google has a very powerful knowledge about the world's dynamics, obtained from the combination of two sources: their web index, which represents a very significant portion of all information available on the WWW, and their search query log, which represents the interest of billions of users on that information. Facebook, on the other hand, seems to have a more limited knowledge about the needs of its users (because they do not request information from it in the same way they do with Google), but it is able to accurately profile its users based on their personal details, their status updates and their "liked" items.

We have also presented a few strategies that Google and Facebook could adopt to make profit from their knowledge, suggesting some examples on how they could make certain investments or act as a consulting company that sells pieces of that knowledge. This would require these Internet-born companies to extend their activities to the off-line world, following the analogous path of many traditional companies that have moved part of their business to the Internet.

The reader must take into account that the market strategies and the discussions presented in this paper are founded in the opinions of the authors, who do not have any professional or academic experience in the business field, but only a background in engineering. Having said that, the paper is not only biased by the authors' lack of expertise in such field, but it is also incomplete, as it merely performs a general overview on the matter and does not cover the whole range of information available to the Internet giants. Further research may involve detailed investigations on the insights of these companies to discover additional bits of gathered data, and also, a more profound analysis of the proposed strategies and their commercial viability.

REFERENCES

[1] Europeean Comission. "CT and e-Business for an Innovative and Sustainable Economy. 7th Synthesis Report of the Sectoral e-Business Watch (2010)". *Unit D4 "ICT for Competitiveness and Innovation". Enterprise and Industry. European Commission.* Europe. 2010. [Available on-line] http://www.ebusiness-watch.org/key_reports/documents/EBR09-10.pdf

[2] Kumar, N. and Ruan, R. "On manufacturers complementing the traditional retail channel with a direct online channel". *Quant Market Econ (2006) 4:289–323. Springer Science + Business Media, LLC*. 2006. [Available on-line] http://www.springerlink.com/content/6021106727442123/

[3] Alexa. "The top 500 sites on the web". *Alexa Internet, Inc. An Amazon.com Company.* United States. (2011). [Available on-line] http://www.alexa.com/topsites

[4] comScore. "comScore Explicit Core Search Share Report: August 2010. U.S. Search Engine Rankings". *comScore, Inc.* United States. (2010). [Available on-line] http://www.comscore.com/Press_Events/Press_Releases/2010/9/comScore_Releases_August_2010_U.S._Search_Engine_Rankings/

[5] YouTube. "YouTube Fact Sheet". *YouTube, LLC.* United States. 2011. [Available on-line] http://www.youtube.com/t/fact_sheet

[6] Facebook. "Facebook Factsheet". *Facebook, Inc.* United States. 2011. [Available on-line] http://www.facebook.com/press/info.php?factsheet

[7] Guynn, J. "Facebook hits 600 million users, executive says". *Los Angeles Times.* Los Angeles, California. United States. January 24, 2011. [Available on-line] http://latimesblogs.latimes.com/technology/2011/01/facebook-hits-600-million-users-facebook-executive-says.html

[8] "Results of Operations and Financial Condition. Consolidated Balance Sheets: Google, Inc. Financial Tables For The Quarter Ended June 30, 2009". *U.S. Securities and Exchange Commission.* United States. 2010. [Available on-line]

http://www.sec.gov/Archives/edgar/data/1288776/000119312509150129/dex992.htm
[9] Chang, F., Dean, J., Ghemawat, S., Silson, Hsieh, WC., Wallach, DA., Burrows, M., Chandra, T., Fikes, A and Gruber, RE. "Bigtable: A Distributed Storage System for Structured Data". *OSDI'06, Seventh Symposium on Operating System Design and Implementation.* Seattle, United States. 2006. [Available on-line] http://labs.google.com/papers/bigtable-osdi06.pdf
[10] AFP. "Social media 'one part' of Google strategy: CFO". *The Independent.* United Kingdom. 23 November 2010. [Available on-line] http://www.independent.co.uk/life-style/gadgets-and-tech/social-media-one-part-of-google-strategy-cfo-2141347.html
[11] IPInfoDB. "IP address geolocation database". *IPInfoDB*. 2010. [Available on-line] http://www.ipinfodb.com/ip_database.php
[12] Google. "Google Latitude". *Google Inc.* United States. 2011. [Available on-line] http://www.google.com/latitude
[13] Facebook. "Facebook Places". *Facebook Inc*. 2011. [Available on-line] http://www.facebook.com/places/
[14] Norvig, P. "Inference In Text Understanding". *AAAI Spring Symposium on Machine Learning. Proceedings of the National Conference on Artificial Intelligence.* p561-565. 2007. [Available on-line] http://norvig.com/aaai87.pdf
[15] Pasca, M. "Turning Web Text and Search Queries into Factual Knowledge : Hierarchical Class Attribute Extraction". *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence.* 2008. [Available on-line] http://www.aaaipress.org/Papers/AAAI/2008/AAAI08-194.pdf
[16] Jiulong , S., Genqing, W., Zhihong, H., Xiliu, T., Jansche, M. and Moreno, PJ. "Search by Voice in Mandarin Chinese". *Google China Engineering Center and Google Research. Interspeech.* 2010. [Available on-line] http://research.google.com/pubs/archive/36463.pdf
[17] Schuster, M. "Speech Recognition for Mobile Devices at Google". *Google Research. Springer-Verlag*. Berlin Heidelberg. 2010. [Available on-line] http://www.springerlink.com/index/L2U4711902061767.pdf
[18] Potamianos, A. and Narayanan, S. "A Review of the Acoustic and Linguistic Properties of Children's Speech". *IEEE 9th Workshop on Multimedia Signal Processing, 2007. MMSP*. 2007. [Available on-line] http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4412809
[19] Masanori, I., Toshio, I., Shigeaki, A. and Tadahisa, K. "An articulatory model with developmental constraints: Some evidence from child speech data". *Proceedings of 2005 4th IEEE International Conference on Development and Learning. IEEE*. 2005. [Available on-line] http://doi.ieeecomputersociety.org/10.1109/DEVLRN.2005.1490945
[20] Revett, K., Tenreiro de Magalhães, S. and Santos, HMD. "On the Use of Rough Sets for User Authentication Via Keystroke Dynamics". EPIA'07 *Proceedings of the artificial intelligence 13th Portuguese conference on Progress in artificial intelligence. Springer-Verlag*. Berlin, Germany. 2007. [Available on-line] http://www.springerlink.com/content/y20x553417242x36/
[21] Zahid, S., Shahzad, M., Khayam, SA. and Farooq, M. "Keystroke-Based User Identification on Smart Phones". *Recent Advances in Intrusion Detection.* Lecture Notes in Computer Science, 2009, Volume 5758/2009, 224-243. 2009. [Available on-line] http://www.springerlink.com/content/j72h5012t87v3423/
[22] He, J., Chu, W. and Liu, Z. "Inferring privacy information from social networks". *Intelligence and Security Informatics.* United States. 2006 [Available on-line] http://www.springerlink.com/content/t81k70nh33k782xh/
[23] Park, SH., Han, SP., Hug, SY. and Lee, H. "Preprocessing Uncertain User Profile Data: Inferring User's Actual Age from Ages of the User's Neighbors". *IEEE 25th International Conference on Data Engineering. ICDE '09*. IEEE. 2009.[Available on-line] http://www.computer.org/portal/web/csdl/doi/10.1109/ICDE.2009.154
[24] Google. "Google Insights for Search". *Google Inc*. United States. 2011. [Available on-line] http://www.google.com/insights/search/
[25] Liao, L., Fox, D. and Kautz, H. "Learning and Inferring Transportation Routines". *Department of Computer Science & Engineering. University of Washington*. United States. 2004. [Available on-line] http://www.cs.rochester.edu/~kautz/papers/gps-tracking.pdf
[26] Broxton, T., Interian, Y., Vaver, J. and Wattenhofer, M. "Catching a Viral Video". *Google Inc*. United States. 2010. [Available on-line] http://research.google.com/pubs/archive/36697.pdf
[27] Hyunyoung Choi, H. and Varian H. "Predicting the Present with Google Trends". *Google Inc*. United States. 2009. [Available on-line] http://www.google.com/googleblogs/pdfs/google_predicting_the_present.pdf
[28] Lerman, K., Gilder, A., Dredze, M. and Pereira, F. "Reading the Markets: Forecasting Public Opinion of Political Candidates by News Analysis". *Conference on Computational Linguistics (Coling).* 2008. [Available on-line] http://research.google.com/pubs/archive/34666.pdf
[29] Fortt, J. "Top 5 moments from Eric Schmidt's talk in Abu Dhabi". *Fortune Tech. CNNMoney.com. Cable News Network*. United States. March 11, 2010. [Available on-line] http://tech.fortune.cnn.com/2010/03/11/top-five-moments-from-eric-schmidts-talk-in-abu-dhabi/
[30] Gartner. "Hype Cycle". *Gartner, Inc*. United States. 2011. [Available on-line] http://www.gartner.com/technology/research/methodologies/hype-cycle.jsp
[31] Atos Origin. "Atos Origin sets out its ambition to be a zero email company within three years". *Global Public Relations Office. Atos Origin.* United Kingdom. 2011. [Available on-line] http://www.atosorigin.com/en-us/Newsroom/en-us/Press_Releases/2011/2011_02_07_01.htm