International Journal of Interactive Multimedia and Artificial Intelligence

March 2014, Vol II, Number 5, ISSN: 1989-1660

Happiness depends upon ourselves

Aristotle (384 BC - 322 BC)

Special Issue on AI Techniques to Evaluate Economics and Happiness

http://www.ijimai.org

International Symposium

DCAI

Distributed Computing and Artificial Intelligence

http://dcai.usal.es

PAAMS. 12th International Conference on Practical Applications of Agents and Multiagent Systems

more information: http://www.paams.net

Program Chairs
- Yves Demazeau - Centre Nat. de la Rech. Scientifique (France)
- Franco Zambonelli - Uni. of Modena and Reggio Emilia (Italy)

Organising committee Chairs
- Juan M. Corchado - University of Salamanca (Spain)
- Javier Bajo - Technical University of Madrid (Spain)

PACBB. 7th International Conference on Practical Applicationsof Computational Biology & Bioinformatics

more information: http://www.pacbb.net

Organising committee
- Florentino Fdez-Riverola - University of Vigo (Spain)
- Juan F. De Paz - University of Salamanca (Spain)

Scientific committee chairs
- Miguel Rocha - University of Minho (Portugal)
- Julio Saez-Rodriguez - European Bioinformatics Institute (Spain)

DCAI. 11th International Symposium Distributed Computing and Artificial Intelligence

more information: http://dcai.usal.es

Organising Committee Chairs
- José M. Molina - Universidad Carlos III de Madrid (Spain)
- James Llinas - State University of New York (USA)
- Andre Ponce de Leon - University of Sao Paulo at Sao Carlos (Brazil)
- Ajith Abraham- Norwegian University of Science and Technology

Scientific Committee Chairs
- Hugues Bersini (Co-Chairman) - Université Libre de Bruxelles (Belgium)
- Sigeru Omatu (Chairman) - Osaka Prefecture University (Japan)

ISAmI. 5th International Symposium on Ambient Intelligence

more information: http://isami.usal.es

Organising Committee Chair
- Juan M. Corchado - University of Salamanca (Spain)
- Paulo Novais - University of Minho (Portugal)

Scientific Committee Chair
- Carlos Ramos - Polytechnic Institute of Porto (Portugal)

SUBMISSION

Review process

We welcome the submission of application papers with preference to the topics listed in the call for papers. All submitted papers will undergo a thorough review process; each paper will be refereed by at least three experts in the field based on relevance, originality, significance, quality and clarity.

Submitting papers

All papers must be formatted according to the Springer template, with a maximum length pages in length (see expecifics conference CFP), including figures and references. All proposed papers must be submitted in electronic form (PDF format) using the Paper Submission Page.

Publication

Accepted papers will be included in the conferences proceeding edited by Springer.

Springer Lecture Notes in Computer Science ccis

IMPORTANT DATES

PAAMS

Submission dates	20th January, 2014
Notification of acceptance	24th February, 2014
Camera-ready deadline	10th March, 2014

DCAI · PACBB · ISAmI

Submission dates	18th January, 2014
Notification date	23rd February, 2014
Paper ready deadline	7th March, 2014

CALL FOR GRANTS/AWARDS

The BISITE research group (http://bisite.usal.es) of the University of Salamanca, organiser of PAAMS 2014, offers 40 grants/awards to facilitate the attendance to doctoral and young doctors both to the event itself (including Workshops and Special Sessions) and to the parallel collocated events.

More information: http://www.paams.net/grants

INTERNATIONAL JOURNAL OF ARTIFICIAL INTELLIGENCE AND INTERACTIVE MULTIMEDIA
ISSN: 1989-1660 - VOL. II, NUMBER 5

Editor's Note

The science of happiness is trans-disciplinary. Happiness is an experience human beings have and, in consequence, its understanding calls for knowledge from all disciplines which, in one way or the other, deal with all facets of human lives. Various disciplines have contributed to the development of the science of happiness; among them: Psychology, Sociology, Economics, Psychiatry and Neuroscience. Because happiness research deals with human being of flesh and blood, it requires high-level techniques to dealing with large information sets in order to extract that information which is relevant. In the study of happiness there are many observations –as many as persons in the world-, there are many variables, and there are many interrelations and synergies to take account of. In consequence, happiness research benefits from sophisticated models that allow for a better understanding of people's happiness; without losing contact with what real human beings experience, it is important to use techniques that allow researchers to process all the information reaching for valuable conclusions. It is with this purpose that Computer Science has joined the other disciplines providing its calculation powerful tools to advance the study of happiness. It is therefore appropriate that The International Journal of Interactive Multimedia and Artificial Intelligence has decided to launch a special issue on happiness showing some of the potential contributions the discipline can make to happiness research.

The research works presented in this issue cover various topics of interest, all related to potential contributions from Computer Science to the understanding of happiness and subjective well-being.

The paper of Yago Saez, Carlos Navarro, Asuncion Mochón and Pedro Isasi, *A System for Personality and Happiness Detection*, proposes a platform for estimating personality and happiness. Numerous studies have been conducted on happiness and personality since Hans Jürgen Eysenck defined in 1947[1] the basic traits that delineate personality. These works have supported the theory of individual differences across humans with regards to personality [2], [3]; most research has used questionnaires as a direct ways to obtain a score for each trait (Psychoticism, Extroversion and Neuroticism) [4]. However, there may be alternative ways to knowing people's personality; one highly-potential way is through analyzing written text. This type of approach offers numerous advantages for researchers because a substantial amount of information about a subject's personality profile can be obtained without their presence or any additional specific effort on the subject's part. Although research on personality profiling and analysis of the written word is part of psychology, collaboration with computer science is necessary to be able to use quantitative methods to analyze large amounts of information [5]. Departing from Eysenck's human personality theory, the authors develop a platform for collecting text messages from social media (Whatsapp), and for classifying them into different personality categories. Although there is not a clear link between personality features and happiness, some correlations between them could be found. In this work, the authors describe the platform developed, and as a proof of concept, they use different sources of messages to see if common machine learning algorithms can be used to classifying different personality features and happiness.

The paper of Francisco Mochón and Oscar Sanjuán, *A First Approach to the Implicit Measurement of Happiness in Latin America through the Use of Social Networks*, aims to measure subjective well-being in Latin America on the basis of large-information databases that require novel information-management techniques [6]. The paper's goal is to verify to what extent two radically different methods are consistent in measuring the happiness of Latin Americans. One method is based on the use of surveys [7], while the other is based on inferring the feelings of social-network users from a semantic analysis of the words used in their communications and messages [8]. Following a previous study by Dodds and Danforth, the researchers have developed a method that, by incorporating a direct assessment of words, allows to measure subjective well-being on a continuous scale from a diverse collection of texts [9]. The method is transparent and able to quickly process texts from the Internet.

The study of happiness requires taking into consideration that it is a living experience that happens to persons and not to individuals [10]. Happiness is experienced by persons who are in society and who are living in their circumstance. Thus, the understanding of happiness requires from incorporating a person's context, which implies for the need of incorporating how people interact with others [11]. The paper of Mariano Rojas and Ignacio Ibarra-López, *Happiness and Human Relations: The Role of Materialistic Values. An ABM Illustration*, argues that a person's happiness must be understood as a phenomenon that emerges not only from her individual condition but also from her place in society [12]. Understanding that a person is socially immersed implies giving a greater role to social interactions and social structure. The paper presents a model to take into consideration the role of human relations. An agent-based model (ABM) is used to illustrate the implementation of the model in understanding people's happiness. This allows stressing the role that social interactions play in generating happiness within different value contexts. In specific, the paper studies how materialistic values influence the way rational agents end up following in their pursuing of happiness [13]. The model recognizes that people do interact in the generation of relational goods and their happiness does not depend on their isolated decisions but also on what their fellows do. The model explains how rational people end up allocating their limited endowment of time between working and relating; the model also assumes that people are statically rational while they are dynamically bounded-rational.

The research conducted by Héctor Cordobés, Antonio Fernández Anta, Luis F. Chiroque, Fernando Pérez, Teófilo

DOI: 10.9781/ijimai.2014.250

Redondo and Agustín Santos deals with *Graph-Based Techniques for Topic Classification of Tweets in Spanish*. / Topic classification of texts is one of the most interesting challenges in Natural Language Processing (NLP) [14]. In the field of the happiness research it is important to combine sentiment analysis with topic classification techniques, in order to determine the reasons why a subject express happiness or sadness. Topic classifiers commonly use a bag-of-words approach, in which the classifier uses (and is trained with) selected terms from the input texts [15], [16]. The work included in this special issue presents techniques based on graph similarity to classify short texts by topic [17], [18]. A prototype classifier was developed and was used to participate in the topic classification challenge of the Workshop on Sentiment Analysis at SEPLN in- 2013. For topic classification, a set of Twitter messages (tweets) in Spanish were provided. The authors build graphs from the input texts, and then use properties of these graphs to classify them.

The work of José Manuel Saiz-Alvarez, Alicia Coduras Martínez and Carlos Cuervo-Arango Martínez, *An Entrepreneurial Well-being Model based on GEM Data for Spain*, focuses in the consideration given to entrepreneurship as a determinant of well-being [19], [20]. The authors present a venture-based model in which satisfaction of Spanish entrepreneurs with their professional life is performed. More specifically the paper studies the relationship between entrepreneurship and well-being using Global Entrepreneurship Monitor (GEM) data. The results show that, for the Spanish case, there is a strong consistency en the results, the opportunity entrepreneurs present greater satisfaction then necessity entrepreneurs

Dr. Francisco Mochón.
Dr. Maríano Rojas.

REFERENCES

[1] Eysenck. H.J, Dimensions of Personality, 1947.

[2] Goldberg, L. R. (1981). Language and individual differences: The search for universals in personality lexicons. In L. Wheeler (Ed.), Review of Personality and Social Psychology, Vol. 2. Beverly Hills, CA: Sage.

[3] Ivanov. A.V, Riccardi. J, A, Sporka. J and Franc. J, "Recognition of Personality Traits from Human Spoken Conversations," Annual Conference of the International Speech Communication Association, 2011.

[4] Eysenck, H.J. and Eysenck, S. B. G., 1985. Revised Eysenck Personality Questionnaire (EPQ-R)

[5] Polzehl. T, Moller. S and Metze.F. Automatically Assessing Acoustic Manifestations of Personality in Speech, 2010.

[6] [6] Kahneman, D., Diener, E. and Schwarz, N. (1999). Foundations of hedonic psychology: scientific perspectives on enjoyment and suffering. New York: Russell Sage Foundation.

[7] Diener, E, Emmons, R.A, Larsem, R.J. and Griffin, S.A..The Satisfaction with Life Scale. Journal of Personality Assessment, 1985, 49, 1, pp 71-75.

[8] Dodds, P. and Danforth, C. (2009). Measuring the Happiness of Large-Scale Written Expression: Songs, Blogs, and Presidents. Journal of Happiness Studies. doi:10.1007/s10902-009-9150-9.

[9] Frank M., Mitchell L., Dodds P., and Danforth C. 2013. Happiness and the Patterns of Life: A Study of Geolocated Tweets. Scientific Reports 2013. Vol. 3, No: 2625, doi:10.1038/srep02625

[10] E. Deci and R. Ryan, Intrinsic Motivation and Self-Determination in Human Behavior, New York: Plenum Press, 1985.

[11] J. Coleman, "Social Capital", in Foundations of Social Theory, J. Coleman, Comp., The Belknap Press of Harvard University Press, pp. 300-324, 1990.

[12] M. Rojas, "Happiness in Mexico: the importance of human relations", in Happiness across Cultures: Views of Happiness and Quality of Life in Non-Western Cultures, H. Selin and G. Davey, Eds., Springer, pp. 241-252, 2012.

[13] L. Blume and S. Durlauf, "Identifying social interactions: a review", in Methods in Social Epidemiology, J. Oakes and J. Kaufman, Eds., Jossey-Bass, San Francisco, 2006.

[14] Mihalcea, R. and P. Tarau. 2004. TextRank: Bringing order into texts. In Proceedings of EMNLP-04and the 2004 Con ference on Empirical Methods in Natural Language Processing, July.

[15] Brin, Sergey and Lawrence Page. 1998. The anatomy of a large-scale hypertextual web search engine. Comput. Netw. ISDN Syst., 30(1-7):107-117, April.

[16] Hassan, Samer, Rada Mihalcea, and Carmen Banea. 2007. Random walk term weighting for improved text classification. International Journal of Semantic Computing, 1(04):421-439.

[17] Blanco, Roi and Christina Lioma. 2012. Graph-based term weighting for information retrieval. Informationretrieval, 15(1):54-92.

[18] Thakkar, Khushboo S, Rajiv V Dharaskar, and MB Chandak. 2010. Graph-based algorithms for text summarization. In Emerging Trends in Engineering and Technology (ICETET), 2010 3rd International Conference on, pages 516-519. IEEE.

[19] Binder, M. & Coad, A. (2013): "Life Satisfaction and Self-employment: A Matching Approach", Small Business Economics, vol. 40, n. 4, pp. 1009-1033.

[20] Lofstrom, M. (2013): "Does self-employment increase the economic well-being of low-skilled workers?", Small Business Economics, vol. 40, n. 4, pp. 933-952.

TABLE OF CONTENTS

OPEN ACCESS JOURNAL

ISSN: 1989-1660

A System for Personality and Happiness Detection

Yago Saez[1], Carlos Navarro[1], Asuncion Mochon[2] and Pedro Isasi[1]

[1]*University Carlos III of Madrid, Computer Science Department, Madrid, Spain*
[2]*Applied Economics Department, UNED, Madrid, Spain*

***Abstract* — This work proposes a platform for estimating personality and happiness. Starting from Eysenck's theory about human's personality, authors seek to provide a platform for collecting text messages from social media (*Whatsapp*), and classifying them into different personality categories. Although there is not a clear link between personality features and happiness, some correlations between them could be found in the future. In this work, we describe the platform developed, and as a proof of concept, we have used different sources of messages to see if common machine learning algorithms can be used for classifying different personality features and happiness.**

***Keywords* — personality detection, Android OS, happiness, written text, machine learning, classifying algorithms**

I. Introduction

Since Hans Jürgen Eysenck in 1947 defined the pillars, or traits, that form personality [1], numerous studies have been conducted and many works have been written about the subject, see Section II. These works have supported his theory of individual differences between humans with regards to personality. This theory is also known as the PEN model because of the three traits on which it is based: Psychoticism, Extroversion and Neuroticism. The theory provides a direct way to obtain a score for each component by using questionnaires, specifically the EPQ-R questionnaire. Each of the three personality traits has a biological basis, so the scores obtained for the traits represent different brain processes.

Researchers have tried to obtain information about the personality of human beings through direct means such as the EPQ-R questionnaire, but they have also used indirect methods. Because personality is considered to be stable over time and throughout different situations, specialized psychologists are able to infer the personality profile of a subject by observing the subject's behavior.

One of the sources of knowledge about the behavior of individuals is written text. According to research in this field, it is reasonable to expect that different individuals will have different ways of expressing themselves through the written word, and these differences will correspond to their individual personality profiles, as well as their moods.

This type of reports offers numerous advantages for researchers because a substantial amount of information about a subject's personality profile can be obtained without their presence or any additional specific effort on the subject's part.

A. Multidisciplinary work

Although research on personality profiling and analysis of the written word is part of psychology, collaboration with other disciplines, such as computer science, is necessary for certain purposes. Even with a solid psychological theoretical foundation, it is also necessary to be able to use quantitative methods to analyze large amounts of information. Such methods are especially applicable when analyzing large amounts of written text.

It is thus necessary to undertake this type of research with a multidisciplinary team, in which social sciences researches and computer scientists combine their knowledge to create efficient tools for the analysis of human personality. Computer science provides the tools necessary to collect, process and classify text samples of psychological interest in a systematic fashion, based on the principles of software engineering and artificial intelligence.

A tool with the aforementioned characteristics will be of great interest for the economy and human happiness. For example, if a system that could recognize the personality traits of a criminal in a matter of minutes with a high degree of confidence was available to law enforcement, a more efficient handling of critical situations could be achieved.

The remainder of this article is structured in the following manner. The following section describes the most relevant works related to this research. Section III describes the objectives and answers to common questions. Section IV depicts Eysenck`s theory of personality background. After Section IV, we describe the proposed platform, (Section V), the classifier module (Section VI) and the preliminary results (section VII). Finally, the main conclusions are presented in Section VIII.

II. State of the Art

The U.S. Army War College has shown an interest in predicting and controlling the behavior of an individual or

DOI: 10.9781/ijimai.2014.251

group of individuals based on knowledge of their personalities. They believe that a system capable of this would have important applications in State security, competition in the labor market, political elections, or simply in the acquisition of knowledge about any person whose behavior might be of interest, see [2].

To perform a strategic personality simulation, they recommend taking into account the intersection between internal and external elements as well as external situational factors and personal influences.

Professors of computer science Gill and Oberlander conducted [3] a study on the recognition of the "extroversion/introversion" personality trait based on written text. They based their work on the Eysenck model [4]. For this purpose, they asked subjects with known scores on the EPQ-R questionnaire to write two e-mails to a fictitious friend. They subsequently analyzed these e-mails with a text analysis program called LIWC (*Linguistic Inquiry and Word Count*) and with the psycho-linguistic database MRC. They generated bigram profiles according to the degree of extroversion of the subjects (high or low). The results showed differences between the two sub-types of samples. Based on these differences, it was found that extroverts use more punctuation and exclamation signs, produce texts with more words, make more references to social situations, and use a greater number of positive words. Introverts, in contrast, are more likely to use the first-person singular, express themselves using more emotionally negative words, and use more coordinating conjunctions. The researchers also made lists of frequently used bigrams for both groups.

With their results, both authors conclude that the personality dimensions have relevance and validity for working with human-computer communication and computer learning.

Young presents in 2003 a geographical profiling, which consists of the profiling of criminals based on questions such as "when" or "where," instead of based on their motivations, age, gender, or other indicators [5]. With this approach, the need to incorporate computer science into the profiling process is emphasized to analyze large databases and prevent people from overlooking important information or connections between crimes. This type of analysis becomes imperative in the case of serial killers, who may commit crimes in different states that involve victims who do not know each other. The proposal coincides with the nature of this project in that it warns about the need for interdisciplinary work and highlights the importance of computer science for the processing of data that individual psychologists would not be able to analyze manually.

In this article [6], the principle of geographic profiling is presented. Geographic profiling is an attempt to obtain a wide body of information about criminal cases to provide a general psychological description of an unknown subject (UNSUB) — a possible suspect. After going into detail about the description of geographic profiling, the author presents several programs for collecting the essential information for this purpose. First, the Violent Criminal Apprehension Program (VICAP) is presented, which is used by the FBI to efficiently analyze the connections between existing criminal cases. Second, Kim Rosso's Criminal Geographic Targeting (CGT) is exhibit. This computer program produces a topographic map by performing many calculations that group together similar crimes, and it takes into account human movement patterns. Lastly, the Predator system, developed by Dr. Grover and M. Godwin, is described. This system uses multivariate analysis to carry out geographic profiling and produces a 3D, color-coded map to classify different areas according to the probability that the perpetrator lives or operates in them.

The word done by F. Mairesse and M. Walker may be considered to be the most important antecedent of the System for Personality Detection (SPD) project [7]. The researchers attempted to automatically identify personalities based on pieces of recorded conversations. Their personality analysis was based on the Five Factor Model (see [8]), which, is closely related to the personality traits of the PEN model used in the present project. In addition to confirming previous studies, the authors reached conclusions about personality. For example, they found that correlations between linguistic indicators and personality traits are higher in informal spoken dialog; this conclusion has stimulated the use of informal language in SPD. They also concluded that the most complex trait to analyze is "neuroticism," whereas "agreeableness" and "conscientiousness" provide the best results. Prosodic indicators were found to be the most accurate predictors for "extroversion." Finally, they concluded that their hypothesis, which proposes that it is possible to automatically detect personality through language, is confirmed, and they find that their procedure is applicable to a variety of fields.

The work of T. Polzehl, S. Moller, and F. Metze shows the results of implementing a personality evaluation paradigm for spoken input, and it compares human and computer performance in carrying out this task [9]. For this investigation, a professional speaker wrote speeches corresponding to different personality profiles, in accordance with the Five Factor Model questionnaire NEO-FFI. Then, human judges who did not know the speaker estimated the five personality factors. Recordings were also analyzed by using methods based on acoustic and prosodic signals. The results were very consistent between the acted personalities (as evaluated by the judges) and the initial classification of the results. Based on this, the authors concluded that they had made a first step toward the use of personality traits in conversations for future human-machine communication.

The study of A. V. Ivanov, G. Riccardi, *et al.* focused on personality prediction in the context of human spoken conversation [10]. For that purpose, once again, the Five Factor Model was used as a reference. The authors' final goal is to create a machine called the *Personable and Intelligent Virtual Agent*, which is capable of adjusting its linguistic behavior as required by the human with whom it converses. This would facilitate human-machine communication. During

this research work, a simulated tourist help agent was created, which gathered linguistic and acoustic information from the subjects taking part is a role-playing game. These individuals volunteered their scores in the Big Five (Five Factor Model) questionnaire, and they were classified by their traits in a binary fashion: high or low. The results showed that machines can be trained to automatically predict personality traits based on conversations. In addition, statistically significant data were presented for the prediction of traits such as "conscientiousness" and "extroversion."

Linguistic Inquiry and Word Count (LIWC) is private software that analyzes text and calculates the degree to which an individual uses words from different categories, see [11]. A wide variety of sources are used, such as e-mails, transcripts of conversations, speeches, and poems. With LIWC, it is possible to obtain, for example, information about the number of emotionally negative words or self-references used, among many other dimensions of language.

Research on the topic of personality is often focused on one trait in particular: extroversion/introversion. Researchers in this field strive to find personality indicators, with the goal of creating simulated human-machine conversations, instead of focusing their discoveries on the creation of tools for personality profiling and happiness analysis. It is worth mentioning that, with the exception of the works [3], [12] and the LWIC2007 package (2007), all investigations were carried out based on spoken conversations and not on written text, in contrast with this work. In any case, existing research focused on the inference of personality and happiness based on the analysis of written text does not make use of mobile devices as a platform.

Regarding the research works that do focus on the creation of profiling tools, they are all centered on geographic profiling; they do not include personality as a factor in the profiling of the subject. Despite this, these works emphasize the need to combine disciplines to produce their tools. That is the spirit of this project.

III. OBJECTIVES

The main goal of this project is to develop a prototype system that is capable to collect information in written Spanish from different sources of interpersonal communication on a mobile device.

The project consists of a module in which a client application is developed for mobile devices running the Android operating system. This application is in charge of compiling and sending information about the user to a server application, which stores the information as it is received.

Independently of the goals set for this work, and according to advances in joint research with a team of criminologists from the Institute of Forensic Sciences and Security (ICFS), work will begin on a prototype for a classifier module that, by processing the collected data, will search for markers to classify the user according to Eysenck's theory of personality.

For this purpose, a system will be created to classify the user based on previously established principles of analysis and natural language processing.

Why mobile devices?

According to a study carried out by CISCO Systems (2013), in 2016, there will be more mobile devices than people, which means that there will be a large number of potential users for the system. In addition, it is worth mentioning that many of the most commonly used means of communication are concentrated on these devices.

Why Android systems?

There are many reasons to implement this project on Android devices, the first of which is that the Android OS provides programmers with more flexibility for the development of applications because it allows for free access to all device resources: an indispensable requirement for the development of the proposed system.

Additionally, the percentage of mobile devices running Android rose to 84.1% by the middle of 2012, according to a study by the consulting company Kantar, i.e., more than four out of five people in Spain who possess a mobile device have one that runs Android. This allows for wider distribution of the application.

Nevertheless, not all Android devices are useful to us, or at least not all of them can provide us with the same sources of information. Because of this, we will focus on smartphones, the devices through which most interpersonal communication takes place.

Why in Spanish?

For the purpose of analyzing the conduct of an individual through their writings, knowing and being able to analyze the language in which the individual expresses himself or herself is paramount, from a psychological point of view. The mere fact that someone uses certain specific words or expressions gives structure to the subject's personality profile. Because of this, a single language must be selected for the development of the application. For the application to be used by people in other countries, it would need to be adapted to the appropriate socio-linguistic context.

This project is being developed in Spain, so the native language (Spanish) of the potential users has been selected.

IV. THEORETICAL BACKGROUND

The theory of personality by Hans J. Eysenck [1] is based on multidimensional taxonomies of personality. From this point of view, there exist personality traits that allow for the description, and therefore prediction, of human personality and conduct, see [13].

Eysenck recognizes three personality traits: psychoticism, extroversion and neuroticism, giving rise to the acronym in PEN theory. These traits manifest themselves in different types of human behavior:

TABLE I
CHARACTERISTICS THAT DEFINE THE THREE PERSONALITY TRAITS OF THE PEN MODEL.

Extroversion	Neuroticism	Psychoticism
Sociable	Irrational	Aggressive
Dominant	Inhibited	Cold
Assertive	Taciturn	Egocentric
Active	Emotional	Impersonal
Lively	Tense	Impulsive
Boastful	Anxious	Antisocial
Daring	Depressed	Creative
Carefree	Feeling guilt	Unfeeling
Adventurous	Low self esteem	Harsh

These traits cannot be understood categorically because they are not mutually exclusive. A subject's personality is composed of three independent traits, which must be understood from a dimensional point of view, [13].

Hence, it is important to understand that the three traits are independent, but together, they determine a personality profile corresponding to the idiosyncrasies of the subject. The potential of their combinations cannot be disregarded.

With this model, an underlying biological basis of the three traits is provided. Eysenck believed that the Extroversion-Introversion trait corresponds to cortical arousal. Specifically, it is controlled by the Ascending Reticular Activating System (ARAS). According to the author, extroverts possess a lower degree of cortical arousal, meaning that they present low cortical activation. In contrast, introverts are *a priori* expected to be highly activated. Given the low "internal" activation of extroverts, they would require external and more intense stimulation, whereas introverts are over-activated and do not require external stimulation to maintain a high level of arousal [14].

The Neuroticism-Stability trait is related to the autonomous nervous system, or the limbic system, which is in charge of regulating emotional impulses. Therefore, a highly neurotic individual will have an unstable autonomous nervous system, leading to intense reactions to stimuli. This would explain the variability of mood and anxiety in neurotic subjects. In stable subjects, the exact opposite would be found, [14].

Psychoticism is the most complicated trait within Eysenck's theory, and only recently has some light been shed on its biological nature. Psychoticism has been found to be related to the vulnerability to psychotic disorders, although this does not mean that people with high scores on this trait are certain to suffer from such personality disorders [14]. The Eysenck Personality Questionnaire-Revised (EPQ-R) [4] is currently used to evaluate the traits proposed by Hans. J. Eysenck.

Lastly, it is worth mentioning the relationship of Eyseck's theory with another multi-trait personality model, which is highly favored by the scientific community: the Five Factor Model. This model, also known as "The Big Five" model [8], is based on five fundamental personality traits: Extroversion, Neuroticism, Openness to Experience, Agreeableness and Conscientiousness [13]. These traits are to be evaluated via the NEO Personality Inventory-Revised (NEO PI-R) [15], or the Big Five Questionnaire (BFQ) [16].

Extroversion and Openness to Experience correspond to the Extroversion trait in PEN theory, Neuroticism has a homologous trait in Eyseck's theory, and Psychoticism would be inversely correlated with Conscientiousness and Agreeableness.

V. TECHNICAL PROPOSAL

In this section, the architecture and design of the system to be developed is presented and the different components of the system are explained.

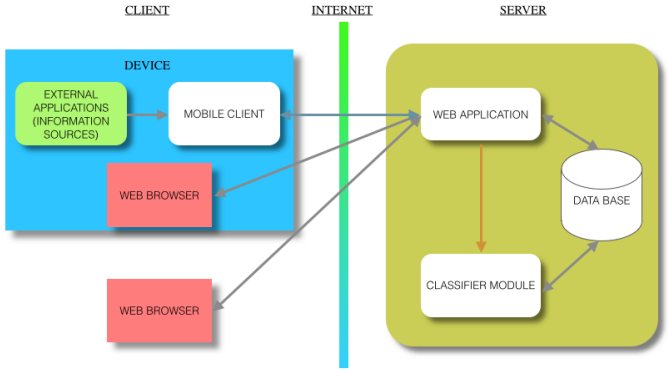


Fig. 1: System architecture

The model to be implemented corresponds to a distributed computer system, which will be composed of numerous devices. Existing classical architectures for distributed systems include the client-server (C/S) architecture and peer-to-peer (P2P) architecture. The C/S architecture is employed when there is a dependency relationship between the devices, which are interconnected in a computer network. This occurs when some functions are performed on the server, and it is the client that communicates with and requests a response from it. In the P2P architecture, every device may function as both client and server.

In the SPD project, there is a logical split within the application. Due to the restrictions described in the non-functional requirements, the system is spread across different computers (physical separation). Only one of the computers—or a group of them functioning as one— will provide services to the rest, thus becoming the "server," the others will submit requests to it, thus becoming "clients." Thus, the chosen architecture is the C/S architecture.

The elements included in the architecture of the SPD system are the following:

- **Client:** software in charge of interacting directly with the user and communicating with the server to submit requests to the system. It will consists of the following:
  - **Mobile device:** the equipment owned by the user, which contains the following elements:
    - External applications: an indispensable aspect of the functioning of the system is that the user has a set of applications for interpersonal communication installed on the device, which will serve as the source of information. The

text samples needed for personality profiling will be obtained from these applications. These sources of information will have to be accessible to the mobile client.

- Mobile client: application that will be developed in this project, which allows the interaction with the user. It will mainly be in charge of gathering information and communicating with the server to classify the personality traits. It will also work as a client to access the information provided by the external applications.

- **Web client:** some actions will have to be carried out through a web client external to the system. It may be located within the mobile device itself, or on any other device with basic web navigation capabilities.

- **Server:** the software to be implemented will run on a computer that is accessible to the clients installed on various user terminals via the Internet. It will itself be in charge of the task of communicating with the clients, providing services such as registration within the system, setting up user accounts, etc. It contains the following elements, which must be differentiated from the server itself:
  - **Web application:** responsible for communicating with both types of clients, mobile and web-based.
  - **Database:** element that will store all user access accounts and all information compiled from the mobile client.
  - **Classifier module:** module that will take care of processing and determining, from the information stored in the database about a given user, which personality profile defines said user.

In addition, both the client and the server will employ a Model-View-Controller (MVC) architecture. This is an architecture typically used for graphical interfaces such as web pages. It separates the components into different layers for the reuse of code, and its decoupling facilitates the development, expansion and maintenance of the application. In this project, a pure implementation of the model will not be used because it will not always be the user who carries out the actions and requests between clients and server.

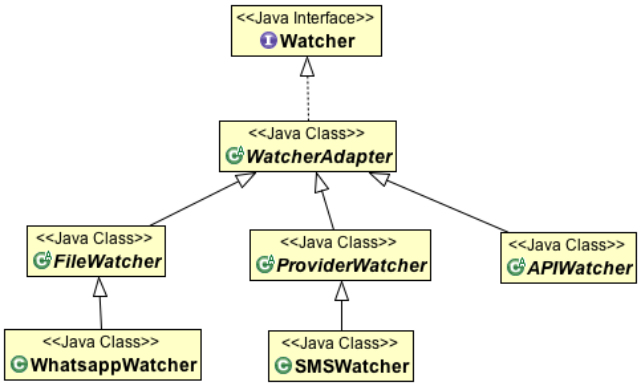


Fig. 2: Details of the hierarchy of implementation of the recollectors.

A. Alternatives

With regard to possible alternatives, we would like to mention that a similar software tool with the same goals as the one presented in this work does not exist, or at least is not publically known. Therefore, this project cannot be implemented using some existing alternative scheme. Notwithstanding, it would be reasonable to carry out an analysis of existing technologies, such as tools or programming languages, to determine which could be useful and to explain why some should be chosen instead of others. We would like to mention that there exists a real-time monitoring software package for Android called *MobileSpy*, which overlaps with SPD as it also collects data. This software, in contrast to the proposed SPD software, collects data from additional sources, such as pictures taken and websites visited by the user. Such information is not necessary for this project, as it is oriented toward espionage.

VI. Classifier Module

In accordance with what was explained in Section III, a prototype classifier module has been programmed as we proceed with the psychological research on the classification and quantification of personality traits through written text.

The classifier is composed of two parts:

First, there is the main process. Its functions include communicating with the database to procure the user data to be processed. In addition, this first part will be in charge of calling the functions of the second part— functions that search for matches of personality indicators— for the evaluation of the established parameters in the compiled user entries. Finally, the main process will have the task of saving in the database the results obtained by the matching functions (for future access, without the need to repeat the analysis on data that have already been processed), and it will be responsible for analyzing the results to evaluate the scores of the different personality traits.

The analysis process is still being developed by the ICFS team and, thus, has not been implemented yet. Basically, they are the search functions that have the task of individually finding and counting every match identified in the data entries, according to the indicators established by the group of psychologists. This group of professionals built a series of lists of indicators to identify personality traits. For example, to identify high scores in the "neuroticism" trait, a list of "emotionally negative words" has been set as an indicator. These indicators are not simply based on the contents of the text, as in the previous example; they also take into account the structure of the text, e.g., lexical density.

Regarding the technical aspects of the classifier, the first decision was to use a language processor to link the analysis to the theoretical guidelines developed by the psychologists. A reference in this field is the *Natural Language Toolkit* (NLTK), but it does not have enough resources to function in the Spanish language. After taking into account many possible solutions and testing a variety of resources to analyze text, we

found that the open source toolkit Freeling was the most approriate for this project. Freeling offers a wide range of functionalities similar to those of NLTK, but with more resources for Spanish. Furthermore, there exists a translation of the WordNet— a lexical database— called the *Multilingual Central Repository* (MCR), which is compatible with *Freeling* and is on par with numerous resources created by other groups. This helped with the automation of the analysis of text using *Freeling*. Thus, the steps needed to implement the language processor have been determined, and this work may proceed as soon as the psychological part of the research allows for it.

VII. First Prototype: Preliminary results

In this work, the aim is to try to classify messages with within the personality features described in Table 1 using supervised machine learning algorithms. The key idea is that these messages have words that will be preprocessed and clustered in order to see whether its possible to match somehow the obtained clusters with the personality features model described before.

However, although the mobile application described before is already developed and ready to deploy we don't have enough information to analyze nor prove the complete method with Whatsapp and SMS messages. There are not many public datasets available in Spanish, so what we have done as a proof of concept is to work with one public dataset with real messages and two datasets generated from messages collected from different websites. For the public dataset we have used a subset of a corpus made of 63.017 *Twitter* messages released by the end of 2013, 17. This dataset has been analyzed for sentiment/opinion analysis with different techniques by several research institutions (IMDEA, LSI-UNED, Elhuyar Fundazioa, L2F, and SINAI-Universidad de Jaen). Other available dataset in Spanish is the 15M Dataset from Complex Systems and Networks Lab (http://cosnet.bifi.es/research-lines/online-social-systems/15m-dataset). This large dataset of *Twitter* messages from the Spanish 15M movement is not useful for this purpose because it is mainly made of hashtags. The other two datasets are made of messages collected manually from several blogs focusing on collecting different lengths of messages (no rights for publishing). For datasets #1 and #2 a classification for extraversion, neuroticism and psychoticism has been made. Personality (based on Eysenck's theory) could be considered an objective measure, but happiness is usually considered opinion. Therefore, there is not a direct translation between the personality features such as extraversion, neuroticism, and psychoticism into happiness. We know that in PEN model, personality it is measured as a weighted combination of features. If written text can express personality features we think that it can express happiness as well, and for future studies we will try to find if any of these combinations can be somehow correlated into an acceptable happiness classification. For these reasons we have decided to include a happiness classification for dataset #3.

Table II
Type of dataset used for testing the clustering algorithms

Dataset	Nº messages	Avg. words / Message	Description
#1	1772	6.53	Twitter
#2	314	10.62	Blogs
#3	200	50.02	Blogs

It is important to remark that the main objective of these experiments is to check if it is possible to identify and classify different types and lengths of messages with common machine learning techniques. In order to do this, we are testing different datasets with several classifying algorithms. First common stage when analyzing texts is to preprocess information. This step will analyze words trying to identify language (English, Spanish, etc.), common lingo, symbols, emoticons, words without semantics and prepositions, articles and conjunctions.

After this initial step, then a stemming analysis takes place. The stemming is a wide used process in information retrieval that reduces or derives words to their stem or root form (if possible). Once this stage finishes, the cleaned dataset is used for clustering.

The first experiments were conducted with WEKA. When working with so many parameters, many different setups can be used. For replication purposes here is a summary of the actions taken, the parameters used and the algorithms tested:

- Preprocessing stage (applied directly to the dataset):
  - Filtering stage (importing data in WEKA):
    - Euclidean distance
    - StringToWordVector filter
    - IDF Transform = True
    - All keywords in lowercase
    - Remove all special characters, punctuation, accents
    - Stopwords: prepositions, articles, conjunctions
    - Eliminate those words that appear less than 5 times.
    - Stemming algorithm: Snowball
    - Deleting keywords with less than 3 letters (before stemming)
    - Deleting numbers
  - Used algorithms (WEKA):
    - J48, binary splits, 0.25 confidence factor
    - Random Forest, 30 trees, no depth limit
    - Support Vector Machine (SVM), LibLINEAR

The Table 3 shows the preliminary results with the three different algorithms and random cross-validation (10 folds).

Table III
Results for all scenarios and algorithms tested

	J48	Random Forest	SVM
#1	Accuracy: 0.42 Kappa: 0.18	Accuracy: 0.57 Kappa: 0.37	Accuracy: 0.56 Kappa: 0.38
#2	Accuracy: 0.65 Kappa: 0.53	Accuracy: 0.58 Kappa: 0.44	Accuracy: 0.65 Kappa: 0.53
#3	Accuracy: 0.64 Kappa: 0.27	Accuracy: 0.64 Kappa: 0.28	Acurracy: 0.82 Kappa: 0.65

The first dataset has in average very short messages for an adequate classification. The accuracy of the tested algorithms in the best case is a 57% (SVM), with many false positives mainly between neuroticism and psychoticism classes (see Table 4 in Appendix). Other issue is the nature of the messages as *Twitter* it is not usually a social media for sending personal messages. Furthermore, most of the selected messages are classified into psychoticism and neuroticism, due to the fact that the dataset comes from primer elections, which bias the type of message collected. These messages are probably really useful for opinion mining but excessively focused on primer elections and short to extract what our psychologists are looking for. The second and third datasets have clearly more promising results. We think that is because messages are longer than the ones from the *Twitter* dataset (see average words per message in Table 2) and because there are not only focused on one specific topic.

VIII. Conclusions

Returning to the goals laid out in Section II, we have created a mobile application for the Android OS that allows for the acquisition of information in the form of written text and allows for communication with a server. In our system the information is acquired from two sources: the *Whatsapp* application and the SMS service. We have also created a server that receives this information from the mobile clients and stores it. Finally, we have also begun work on the creation of a classifier for the gathered information. A prototype has been developed that will continue to be improved.

With regard to the objectives established previously, it can be concluded that all the requirements have been successfully met, except for one: the requirement for the presentation of results in the mobile application. To meet this requirement, it will be necessary to fully develop the classifier and the prototype but this work is still in an early stage although it is a great starting point for future analysis. A lot of work needs to be done prior having more conclusive results. However, so far, we are proud to say that the first objective is achieved. A lot of time has been needed for collaborating with psychologists, understanding the possible ways of identifying personalities and their link with happiness. Finally, developing an application able to collect the *Whatsapp* messages has been a hard task but will allow us the possibility of building the first public corpus of *Whatsapp* messages up to the date. The future of this dataset depends on the overcoming of other problem encountered: the reticence of users to install the application and their non-objective use when installed.

A. Future lines of work

In the short term, we have several goals for continuing the work already done in this project.

First of all, we consider very important to be able to carry out an initial research, so the preliminary results can guide our team and the psychologists to refine the whole system. This first step is already set to be carried out in the near future, the application will be installed on the mobile phones of experimental subjects who will have taken the EPQ-R personality test beforehand. After that, information will be collected and classified by using the tools that have already been developed, which by then will have been improved. The appropriate statistical analysis will be carried out to compare the results of the personality test with the application information. With these results, we will be able to determine how fine-tuned the classifier and the theoretical psychological guidelines are, and we will see what aspects need to be improved so that the personality profile obtained from the application can be trusted. Additionally, we may introduce Artificial Intelligence techniques, such as genetic algorithms, neural networks or automatic learning techniques, as they may help improve the classification.

As mentioned in the previous section, it will be necessary to complete and improve the classifier; this task is intimately related to the research to be carried out. The classifier is a part of the system that cannot be considered finished with only a first build. It will be necessary to verify the proper functioning of the global system and to perform certain adjustments in response to the results. Hence, the system will be improved due to advancements in psychological practice as well as future research that may be carried out.

We are determined to refactor this application and its principles beyond the criminological and clinical fields and to extend the technology to personal use. The field in which we are very interested is human-machine communication. If a machine were able to automatically identify personality from the text provided by a person, it could adapt to the needs and tastes of the person. As [18] mentioned, the moment we are able to identify personality, we are one step closer to being able to simulate it. That is, we could be closer to achieve artificial intelligence.

IX. Acknowledgments

The authors wish to acknowledge Álvaro Ortigosa Juárez (CEO at ACC -Agencia de Certificaciones de Ciberseguridad – and CNEC – Centro Nacional de Excelencia en Ciberseguridad), Manuel de Juan Espinosa (CEO at ICFS – Instituto de Ciencias Forenses y de la Seguridad), Carlota Urruela Cortés (project leader), Irene Gilpérez López and Pilar González Villasante (Main Psychology Team at the ICFS) for their support and collaboration during an important part of this work.

X. Appendix

Detail of confusion matrix obtained with different algorithms and datasets.

TABLE IV
CONFUSION MATRIX FOR TESTS DONE WITH DATASET #1

J48 Confusion Matrix: Dataset #1

Act \ Pred	Extraversion	Neuroticism	Psychoticism	Unknown
Extraversion	**62**	184	48	18
Neuroticism	85	**354**	150	48
Psychoticism	70	138	**247**	115
Unknown	25	19	129	**80**

Random Forest Confusion Matrix: Dataset #1

Act \ Pred	Extraversion	Neuroticism	Psychoticism	Unknown
Extraversion	**114**	187	10	1
Neuroticism	46	**507**	82	2
Psychoticism	1	192	**365**	12
Unknown	1	13	210	**29**

SVM Confusion Matrix: Dataset #1

Act \ Pred	Extraversion	Neuroticism	Psychoticism	Unknown
Extraversion	**199**	102	9	2
Neuroticism	130	**394**	106	7
Psychoticism	22	162	**312**	74
Unknown	1	14	147	**91**

TABLE V
CONFUSION MATRIX FOR TESTS DONE WITH DATASET #2

J48 Confusion Matrix: Dataset #2

Act \ Pred	Extraversion	Neuroticism	Psychoticism	Unknown
Extraversion	**51**	7	22	0
Neuroticism	8	**47**	8	13
Psychoticism	30	12	**30**	7
Unknown	12	14	8	**45**

Random Forest Confusion Matrix: Dataset #2

Act \ Pred	Extraversion	Neuroticism	Psychoticism	Unknown
Extraversion	**52**	6	17	5
Neuroticism	0	**58**	5	13
Psychoticism	32	16	**26**	5
Unknown	9	18	5	**47**

SVM Confusion Matrix: Dataset #2

Act \ Pred	Extraversion	Neuroticism	Psychoticism	Unknown
Extraversion	**49**	3	21	7
Neuroticism	0	**56**	5	15
Psychoticism	16	10	**43**	10
Unknown	2	16	5	**56**

TABLE VI
CONFUSION MATRIX FOR TESTS DONE WITH DATASET #3

J48 Confusion Matrix: Dataset #3

Act \ Pred	Happy	Unhappy
Happy	**670**	330
Unhappy	403	**597**

Random Forest Confusion Matrix: Dataset #3

Act \ Pred	Happy	Unhappy
Happy	**702**	298
Unhappy	417	**583**

SVM Confusion Matrix: Dataset #3

Act \ Pred	Happy	Unhappy
Happy	**844**	156
Unhappy	193	**807**

REFERENCES

[1] H. J. Eysenck, Dimensions of Personality, 1947.

[2] N.D. Livergood, Strategic Personality Simulation: A New Strategic Concept. U.S. Army War College, 1995

[3] Gill, A. and Oberlander, J. [2002] Taking care of the linguistic features of extraversion. In Proceedings of the 24th Annual Conference of the Cognitive Science Society, pp363--368. Washington DC, August 2002

[4] Eysenck, H.J. and Eysenck, S. B. G., 1985. Revised Eysenck Personality Questionnaire (EPQ-R)

[5] G. Young, "Mapping Mayhem: The Geography of Crime," Computoredge, 2003.

[6] K. Rossmo, Profiling and Geography, ISBN-0849381290, 1999.

[7] F. Mairesse, M. and Walker, M., 2006. Automatic Recognition of Personality in Conversation. In: Association for Computational Linguistics ed., Proceedings of the Human Language Technology Conference of the North American New York: Association for Computational Linguistics. Automatic Recognition of Personality in Conversation, pp. 85

[8] Goldberg, L. R. (1981). Language and individual differences: The search for universals in personality lexicons. In L. Wheeler (Ed.), Review of Personality and Social Psychology, Vol. 2. Beverly Hills, CA: Sage.

[9] T. Polzehl, S. Moller and F. Metze, Automatically Assessing Acoustic Manifestations of Personality in Speech, 2010.

[10] A.V. Ivanov, G. Riccardi, A.J. Sporka and J. Franc, "Recognition of Personality Traits from Human Spoken Conversations," Annual Conference of the International Speech Communication Association, 2011.

[11] J. W. Pennebaker, R.J. Booth and M.E. Francis, Linguistic Inquiry and Word Count (LIWC2007), 2007.

[12] M. de Juan-Espinosa and A. Ortigosa Juárez, Detección Automática De La Personalidad En Comunicación Vía Ordenador. Madrid: Universidad Autónoma de Madrid, n.d.

[13] B. Moreno, Psicología De La Personalidad. Procesos. International Thomson Editores, 2008.

[14] M. de Juan-Espinosa, Personalidad y Criminología, Apuntes De Psicología Criminológica. Madrid: Universidad Autónoma de Madrid, 2005.

[15] Costa, P.T. and McRae, R.R., 1990. Revised NEO Personality Inventory (NEO PI-R).

[16] Caprara, G.V., Barbarranelli, C. and Borgogni, L., 1993. Big Five Questionnaire (BFQ).

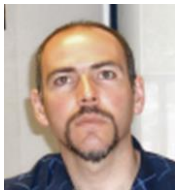
[17] Villena-Román, J., García-Morera, J., Moreno-García, C., Ferrer-Ureña, L., Lana-Serrano, S., Carlos González-Cristóbal, J. Westerski, A., Martínez-Cámara, E., García-Cumbreras, M.A., Martín-Valdivia, M.T., Alfonso Ureña-López, L., 2012, TASS-Workshop on Sentiment Analysis at SEPLN, Workshop on Sentiment Analysis, Sociedad Española para el Procesamiento del Lenguaje.

[18] M. de Juan-Espinosa, Personalidad Artificial: Hacia Una Simulación De Las Diferencias De Personalidad En Situaciones De Interacción. Madrid: Universidad Autónoma de Madrid, 1997.

Yago Sáez received the degree in computer engineering in 1999. He got his PhD in Computer Science (Software Engineering) from the Universidad Politécnica de Madrid, Spain, in 2005. He is vice-head of the Computer Science Department from the Carlos III University of Madrid since 2008, where he got his tenure and is associate professor. He belongs to the Evolutionary Computation, Neural Networks and Artificial Intelligence research group (EVANNAI) and takes part of artificial neural network excellence association (REDAF). He also belongs to the Computational Finance and Economics Technical Committee of the IEEE Computational Intelligence Society. He is involved in several competitive national and international research projects in collaboration with public and private institutions. His main research areas encompasses the evolutionary computation, the computational economic and finance applications and the optimization by means of metaheuristics. Since 2012 he has a very active relationship with Stanford University related to auctions.

Carlos Navarro received the degree in computer engineering in 2013 from the Universidad Carlos III de Madrid, Spain. He has worked in IT companies for three years. Since 2012 he has also has been working as a collaborator in the ICFS (Institute of Forensic Science and Security). Now he is studying a Master Degree in Research and Innovation in Information and Communication Technologies and a Master Degree in Computer Engineering in the Universidad Autónoma de Madrid.

Dr. Asuncion Mochon is currently associate professor at the Department of Applied Economics and Economic History at UNED, Madrid, Spain. She received her Ph.D. in Economics with Honors from UNED University in 2003 and her degree in Economics from the Universidad Autonoma de Madrid in 1998. Her research has focused on several topics such as computational economics, auctions and happiness. She has written numerous articles and participated in multiples conferences on these topics. Most of her recent work, both theory and practice, has addressed questions in economics by means of evolutionary computation techniques and game theory analysis.



Pedro Isasi: Graduate and Doctor in Computer science by the Polytechnical University of Madrid since 1994. Actually he is University professor and head of the Neural Networks and Evolutionary Computation group of the department of Computer Sciences in the Carlos III of Madrid University. Professor Isasi has been Chair of the Computational Finance and Economics Technical Committee (CFETC) of the IEEE Computational Intelligence Society (CIS). His research is centered in the field of the artificial intelligence, focusing on problems of Classification, Optimization and Machine Learning, fundamentally in Evolutionary Systems, Metaheuristics and artificial neural networks.

A First Approach to the Implicit Measurement of Happiness in Latin America Through the Use of Social Networks

Francisco Mochón[1], Oscar Sanjuán[2]

[1]*Universidad Nacional de Educación a Distancia, Madrid, Spain*
[2]*Universidad Carlos III de Madrid, Spain*

***Abstract* - This research paper can be classified as pertaining to the group of empirical studies that try to measure subjective well-being. The article presents as its greatest contributions the use of a subjective measurement of well-being based on social networks for the Latin American setting, as well as its comparative analysis with another traditional method.**

***Keywords* - Happiness, subjective well-being, social networks, Twitter, Latin America, Latinobarómetro.**

I. INTRODUCTION

THIS research paper can be classified as pertaining to the group of empirical studies that for some years now have attempted to analyze subjective well-being in Latin America. Among them some of the most noteworthy are [15], [16], [17], [18], [28], [29], [30] y [5].

The novelty of this paper with respect to previous studies is that its objective is to verify to what extent the results of measuring the happiness of Latin Americans obtained following two radically different methods are consistent. One is based on the use of surveys from Latinobarómetro and the other on inferring the feelings of social network users from a semantic analysis of the words used in their communications and messages. A scientific method is followed in both cases.

The scientific study of happiness is not based on conjectures or presumptions but instead on research projects. Traditionally, researchers have analyzed factors that influence whether an individual defines herself as happy or satisfied [3] [4]. Psychology, sociology and economics have tried to explain the conditions that allow individuals to develop as happy persons [20], [13], [14] and [30].

Following [9][10], the notion of happiness generally used in economics identifies happiness and developing subjective well-being. In this sense, happiness or subjective well-being is no more than an assessment of life itself, regardless of pyschological judgments about momentary pleasure [2], [27]. In other words, happiness refers to how the individual evaluates the overall quality of her life [26], [7]. As such, the happiness of individuals will depend entirely on an individual perception and it will be linked to concepts of quality of life and well-being. In any case, what matters is that that individual perception about the state of subjective well-being or happiness is measurable. This is the notion of happiness that we will use in the third epigraph of the paper, where we use data from Latinobarómetro to measure the happiness of Latin Americans. In the fourth epigraph, we take a completely different look at happiness, using information contained in messages sent over social networks—in particular, data from Twitter--to infer the feelings of individuals [8], [12].

The paper is organized as follows. In this section, we present a short introduction to the study. The following section gives a brief description of the different methods used to measure happiness. As already mentioned, Sections 3 and 4 present two alternative measurements of Latin Americans' happiness, one based on the information gathered from subjective surveys and the other inferred on the basis of information contained in social networks. The fifth section presents a comparative analysis of the results obtained following the two aforementioned methods. The sixth section presents the main findings and sketches out lines of future research that will be conducted to more deeply explore the subjects presented in this paper.

II. THE MEASUREMENT OF HAPPINESS: ALTERNATIVE METHODS

Happiness is measurable, and this is what enables us to speak of the science of happiness. In the new science of happiness, different methods have been used to measure happiness. Ed Diener and his collaborators presented a method to measure happiness based on the idea that individuals can consistently identify their level of satisfaction with life on a scale, and as such, what must be done is to ask people questions [7]. This way of measuring happiness is the one that justifies conducting surveys like the World Values Survey, and it is the most widely-used method [23].

Another method for measuring happiness is based on the sampling of experiences developed by the psychologist

DOI: 10.9781/ijimai.2014.252

Csikszentmihalyi and several researchers. This method consists of using locators (beepers) and afterwards using computers to contact individuals at random and ask them about their mood [24], [25].

A different approach is followed by a group of researchers led by Nobel Prize winner Daniel Kahneman. They created a method for measuring happiness based on following or reconstructing what people do at each moment of the day and asking them how they feel [19]. The main findings of this research specify that the three basic components of happiness are pleasure, commitment and meaning. Following this method and using messages on mobile telephones as an instrument of communication with those surveyed, Matthew Killingsworth identified happiness associated with a wide range of activities [22]. He points out that, until recently, researchers had to trust the assessments and appraisals that people made about their average emotional states over long periods of time. This inconvenience is avoided when following the method based on reconstructing what people do at different moments every day.

Recently, and amidst the impressive growth of social networks, there has emerged a new method for measuring happiness. This method consists of inferring the feelings of social network users on the basis of a semantic analysis of the words used in their communications and messages. Likewise, a study done by the Vermont Complex Systems Center uses information from Twitter to infer how happy or unhappy people in different states of the United States feel. Specifically, the researchers Dodds and Danforth have developed a method that, by incorporating the direct human evaluation of words, allows us to quantify levels of happiness on a continuous scale from a diverse collection of texts [8] [12], [21]. The method is transparent and able to quickly process texts from the Internet.

In the study carried out by Dodds and Danforth, on the basis of ten million "tweets," a code for determining to what extent each analyzed message can be catalogued as happy or sad was developed. The study focused on certain key words that were deemed to be indicative. Thus, "beauty" and "hope" are associated with happiness, while "hate" and "smoke" are associated with unhappiness. The researchers analyze the frequency with which the identifying words are used as good words and bad words in different states of the U.S.A. and qualify them as happy or unhappy. It is important to note that this study requires a highly complex task beforehand that allows us to obtain the terms to evaluate, that is, the words susceptible to be captured and measured. This list of words was obtained by directly asking English-speaking people about the words that evoke happiness for them. Once the list of words was obtained, it was then necessary to create a scale that reflected how one word was evaluated with respect to the following one. This scale was obtained through a similar method, asking people to order words according to the value in terms of happiness that each word had for each of them..

III. The Happiness of Latin Americans according to Latinobarómetro

The measurement of happiness that this part of the study presents is in keeping with the literature that analyzes the answers of individuals to questions about subjective well-being in cross-section or panel surveys, and which is the most widely-used by researchers. The hypothesis on which these studies are based is that the subjective data provided by individuals can be treated ordinally in economic analyses so that greater subjective levels of well-being reflect greater levels of happiness [13]. In other words, it is argued that although everybody has their own ideas about happiness, individual happiness can be captured and analyzed.

Anyone can be asked how satisfied they feel with the life they lead, and behind the answer given in a survey, a conscious evaluation of their subjective well-being can be found. Supposedly, individuals are able to evaluate their subjective level of well-being with respect to certain circumstances. In addition, reliable studies indicate that the subjective well-being demonstrated by individuals is reasonably stable and sensitive to changes in circumstances. In fact, in research about happiness, individuals' answers to questions about their feelings are analyzed and consistent findings are obtained [6].

Specifically, in this section of the paper, there is a synthesis of a paper done in 2012 on life satisfaction in 18 Latin American countries [5]. The countries analyzed are Argentina, Bolivia, Brazil, Colombia, Costa Rica, Chile, Ecuador, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Panama, Paraguay, Peru, the Dominican Republic, Uruguay and Venezuela. The results obtained in the cited study are in general consistent with those already known for other countries, as well as with those obtained in different papers that refer to the region.

The data used come from annual personal surveys created by the Latinobarómetro Corporation for the period 2000-2009. The sample used includes 191,488 individuals and they are different every year. The distribution of the sample by countries over the period is presented in Graph 1, where the information about Brazil has been omitted, given that in the study based on social networks for this country, "tweets" were not analyzed because they were in a different language.

The key variable is the degree of satisfaction with individuals' current lives, as it is defined in the Latinobarómetro survey. The degree of a person's satisfaction with life falls into one of the following four categories: not at all satisfied, not satisfied much, quite satisfied and very satisfied. Graph 1 presents the percentage of individuals from the Latin American countries mentioned that indicate they were quite or very satisfied with life during the years 2000-2009. As can be seen, in eight of the 17 countries, more than 70% of the population was quite or very satisfied with their life at the time. Peru, Bolivia and Ecuador are the countries where people were least satisfied with life. In these countries, it can be seen that less than 54% of the people surveyed indicated that they were quite or very satisfied with life.

GRAPH 1

PERCENTAGE OF THE LATIN AMERICAN POPULATION THAT WAS QUITE OR VERY SATISFIED OVER THE PERIOD 2000-2009

Peru	43.56
Bolivia	48.63
Ecuador	53.15
Nicaragua	61.35
El Salvador	65.02
Chile	65.03
Argentina	66.94
Paraguay	67.65
Honduras	68.15
The Dominican Republic	70.14
Uruguay	71.54
Mexico	72.07
Guatemala	73.97
Colombia	74.13
Panama	76.69
Venezuela	78.81
Costa Rica	81.32

Source: De Juan and Mochón (2012)

A more rigorous analysis of the happiness of Latin Americans is obtained by studying the satisfaction with life variable by countries. The results show that there are significant differences in the average level of satisfaction with life between the countries studied (Table 1). Of the group of countries analyzed, only eight show a coefficient of satisfaction with life higher than 2 (quite satisfied). These are Costa Rica (2.234), Venezuela (2.173), Panama (2.086), Colombia (2.068), the Dominican Republic (2.035), Guatemala (2.034), Honduras (2.022) and Mexico (2.010). Also, there are six countries that have an intermediate value, between 1.76 and 1.89. They are El Salvador (1.891), Uruguay (1.880), Paraguay (1.844), Nicaragua (1.835), Argentina (1.785) and Chile (1.764). The countries that show a lower coefficient are Ecuador (1.614), Bolivia (1.519) and Peru (1.484).

TABLE 1.
AVERAGE LEVEL OF SATISFACTION WITH LIFE

Countries	Satisfaction with life
Costa Rica	2.234
Venezuela	2.173
Panama	2.086
Colombia	2.068
Dominican Republic	2.035
Guatemala	2.034
Honduras	2.022
Mexico	2.01
El Salvador	1.891
Uruguay	1.88
Paraguay	1.844
Nicaragua	1.835
Argentina	1.786
Chile	1.764
Ecuador	1.614
Bolivia	1.519
Peru	1.484

Source: De Juán and Mochón (2012)

So, from the descriptive analysis carried out on the basis of the information provided by Latinobarómetro, it can be seen there are significant differences between countries in the level of satisfaction with life. These results indicate that the happiest individuals are those who live in Costa Rica, Venezuela, Panama and Colombia, while the least happy are those in Peru, Bolivia and Ecuador. It must be noted that these results are consistent with the results obtained in [5] using econometric techniques.

IV. THE HAPPINESS OF LATIN AMERICANS ACCORDING TO SOCIAL NETWORKS

The boom that social networks are currently experiencing is well-known, and their great reach justifies their use as a medium to measure opinion, interest in a subject or a person or even feelings and moods[31] [32] [11].

Especially relevant is the use of social networks in marketing and publicity, the measurement of audiences, opinion surveys, popularity and even as previews of election results. Resorting to social networks to obtain a barometer of opinion is especially common in the social network Twitter, where, since its creation, it has been possible to know the number of followers or the effect of a speceific term or tag [8]. Keep in mind that there are also tools that facilitate more rigorous analysis and establishing relationships, measuring impacts, etc.

To summarize, the reasons that can justify choosing Twitter as a tool for measuring interest, opinion or mood are as follows:

1. Availability of an API (Application Program Interface): the existence of a public API makes it possible to make consultations and recover information in a relatively simple way, through the creation of simple computer programs that facilitate recovery, storage and analysis using different techniques ranging from basic statistics to machine learning.
2. Simple content based on text: the most usual type of message on Twitter is the short text message, owing to its origin from when messages were sent and received via SMS. This characteristic requires the meaning of the messages to be direct, specific and simple in most cases, which helps in their analysis.
3. Instantaneity and transience: the instantaneity and simplicity of the messages on Twitter make it a good

mechanism for measuring what happens in almost real time or during a period of time. They are not reflexive, prepared publications but spontaneous, fast communications.

4. Profiling: many of the users on Twitter not only make comments but also have a public profile, which allows for their segmentation according to this data.
5. Geographical segmentation: within any mechanism for measuring opinion, a basic factor is knowing where we are measuring. On Twitter, this is possible through both the user profile and the location of a specific publication.
6. Global use: although Twitter is not the most used social network, it has many users and a very high level of participation [12].

The use of Twitter to measure subjective well-being as presented in this study is not completely new. As has been noted, there is a project called "hedometer" (http://hedonometer.org) that has taken a measurement of happiness (subjective well-being) in the United States of America [8]. This measurement is especially interesting as it demonstrates the possible use of social networks to measure happiness. In addition, it has other interesting characteristics, like being able to take the measurement in a large geographical area with a common language, and being a space where the use of social networks in general and Twitter in particular is very widespread.

Taking these characteristics as a framework of reference, a similar study has been undertaken in our case, in another relatively homogeneous geographical environment and in a common language. Specifically, the study was carried out for the Spanish-speaking countries of Latin America. Although a study with these characteristics can be valuable in itself, it was interesting to contrast the results with the results obtained when a traditional method of measuring happiness is used.

To produce the present research paper, the following considerations have been taken into account:

1. The recovery of "tweets" for a national geographical area is very complex and unreliable, since it can only be based on the data in the personal profile of each user, and this information is not usually contributed by the users. This is why we have decided to use the "tweets" recovered from the capitals of each country as a representative sample to analyze. This process is rather more simple than if we try to use the personal profile of each user, and more effective since Twitter allows consultations which indicate a geographical position and a sphere of interest.
2. The recovery of "tweets" has been done for a group of key words obtained, taking as a reference the group of key words that hedometer uses [8]. These key words are logically in English, which is why they have been translated. As we are dealing with key words, the idiomatic and semantic problems of translation can be managed. In any case, we have eliminated those that could present some problem. Obtaining this list (Table 2) has a certain value, since it was created on the basis of a thesaurus, considering the different words according to their meaning and impact as indicators of happiness based on the information provided by [8]. A thorough process of translation was applied to the original list, eliminating those words that make no sense in Spanish.

TABLE 2. LIST OF WORDS AND WEIGHTS

WORD	VALUE	WORD	VALUE	WORD	VALUE	WORD	VALUE	WORD	VALUE	WORD	VALUE	WORD	VALUE	WORD	VALUE	WORD	VALUE
carcajada	8,5	fin de semana	8	vacacion	7,92	ganador	7,78	bonus	7,68	mono	7,62	victorias	7,58	comodidad	7,5	leal	7,46
felicidad	8,44	celebrar	7,98	mariposas	7,92	delicia	7,78	brillante	7,68	entretenimiento	7,62	conseguido	7,56	felicitaciones	7,5	oportunidades	7,46
amor	8,42	comedia	7,98	libertad	7,9	belleza	7,76	diamantes	7,68	excitado	7,62	billón	7,56	magdalena	7,5	triunfo	7,46
feliz	8,3	bromas	7,98	flor	7,88	mariposa	7,76	dia libre	7,68	excitación	7,62	tartas	7,56	ganar	7,5	guau	7,46
reido	8,26	rico	7,98	grande	7,88	entretenimiento	7,76	suerte	7,68	broma	7,62	facilisimo	7,56	extraordinario	7,5	joyas	7,46
risa	8,22	victoria	7,98	luz solar	7,88	el más divertido	7,76	madre	7,68	millonario	7,62	flores	7,56	gloria	7,5	bosques	7,45
riendo	8,2	navidad	7,96	amorcito	7,88	honestidad	7,76	super	7,68	premio	7,62	regalos	7,56	gracioso	7,5	manzana	7,44
excelente	8,18	libre	7,96	dulcito	7,88	cielo	7,76	increible	7,66	consiguió	7,62	oro	7,56	luz de luna	7,5	sueños	7,44
risas	8,18	amistad	7,96	premio	7,86	sonrisas	7,76	angeles	7,66	exitosamente	7,62	mirra	7,56	optimista	7,5	fantasía	7,44
disfruto	8,16	diversion	7,96	chocolate	7,86	conseguido	7,76	disfrutar	7,66	ganadores	7,62	familias	7,54	en paz	7,5	comida	7,44
exitoso	8,16	vacaciones	7,96	jajajaja	7,86	maravilloso	7,76	amigo	7,66	brillas	7,6	guapo	7,54	romance	7,5	miel	7,44
ganar	8,12	amado	7,96	cielo	7,86	glorioso	7,74	amistoso	7,66	fenómeno	7,6	amantes	7,54	festivo	7,49	milagros	7,44
arcoiris	8,1	amados	7,96	paz	7,86	besos	7,74	madre	7,66	genio	7,6	afecto	7,53	atractivo	7,48	sexo	7,44
sonrió	8,1	amando	7,96	espléndido	7,86	promoción	7,74	beneficio	7,66	logro	7,58	caramelo	7,52	contento	7,48	cantar	7,44
gane	8,1	playa	7,94	exitoso	7,86	familia	7,72	mejor	7,66	tarta	7,58	tierno	7,52	abuelita	7,48	luz de las estrellas	7,44
placer	8,08	jajaja	7,94	disfrutando	7,84	regalo	7,72	mal día	7,64	brindemos	7,58	diamante	7,52	internet	7,48	agradecido	7,44
sonríe	8,08	besando	7,94	besado	7,84	humor	7,72	campeón	7,64	excitante	7,58	ganancias	7,52	agradable	7,48	gané	7,44
arcoiris	8,06	sunshine	7,94	atracción	7,82	romantico	7,72	abuela	7,64	bondad	7,58	interesante	7,52	ganancias	7,48	logro	7,42
ganando	8,04	hermoso	7,92	celebrado	7,8	magdalenas	7,7	jaja	7,64	abrazo	7,58	pac_ficamente	7,52	listo	7,48	adorado	7,42
celebración	8,02	delicioso	7,92	heroe	7,8	festival	7,7	beso	7,64	ingresos	7,58	piropo	7,52	navidad	7,48	afectivo	7,42
disfrutó	8,02	amigos	7,92	abrazos	7,8	jajajajaja	7,7	gatito	7,64	fiesta	7,58	rosas	7,52	bebés	7,46	afecto	7,42
saludable	8,02	divertido	7,92	positivo	7,8	honor	7,7	milagro	7,64	sonriendo	7,58	sábados	7,52	brindo	7,46	la vida es bella	7,42
música	8,02	sobresaliente	7,92	sol	7,8	relax	7,7	dulce	7,64	canción	7,58	fiel	7,51	coraje	7,46	esto es vida	7,42
celebrando	8	paraiso	7,92	cumpleaños	7,78	angel	7,68	bendiciones	7,62	éxito	7,58	cielos	7,51	entusiasmo	7,46	vivo	7,42
felicitaciones	8	dulcisimo	7,92	fantástico	7,78	mal día	7,68	brillo	7,62	sabor	7,58	apreciar	7,5	honesto	7,46		

Source: compiled by the authors from the list using in the "hedometer".

3. The use of Twitter to make the ranking of the different countries according to inferred happiness on the basis of the contents of "tweets" has the problem of showing a strong dependence on the intensity or frequency of Twitter use in each country. The creation of a coefficient was produced by taking into account the studies that are compiled in the source in Table 3 as a reference.

TABLE 3. LIST OF CAPITALS AND COEFFICIENTS OF TWITTER USE

City	Use Ratio
Caracas	0.276
Bogota	0.184
Montevideo	0.123
Buenos Aires	0.191
Mexico	0.1616
Santiago	0.156
Asunción	0.123
Guatemala	0.12
Lima	0.1
Quito	0.05
La Paz	0.05
Santo Domingo	0.05
Panama	0.05
San Jose	0.05
Managua	0.05
Tegucigalpa	0.05

Source: This coefficient is the result of the information referring to the percentage of Twitter use in different countries and its creation is based on what was done at: http://alt1040.com/2011/04/los-10-paises-mas-adictos-a-twitter; The Netherlands Ranks #1 Worldwide in Penetration for Twitter and LinkedIn http://bit.ly/1fh0Ql8; and Twitter Grows Stronger in Mexico - eMarketer http://po.st/1WR61k. For those places that did not have a reference value, we assigned 0.5 percent.

Once the plan for carrying out the study was established (how the data would be obtained and under what conditions), we proceeded to design an algorithm to extract the information. The extraction algorithm was executed on Twitter for the duration of the study in order to obtain a happiness ranking. The extraction and generation of the happiness ranking was done according to the following process:

1. For each country (City) on the list:
    a. For each word on the list:
    i. Recover the corresponding "tweets"
    b. They are added up
    c. The happiness factor of the word is applied
2. A total is obtained
3. The correction of Twitter use is applied
4. The list of countries is ordered according to the score obtained
5. The ranking is generated

This process was executed on Twitter for two months to obtain a sample size large enough to be able to obtain significant results. The number of "tweets" used was 100,000.

V. HAPPINESS IN LATIN AMERICA ACCORDING TO TWITTER

As has been mentioned, to be able to apply the algorithm described in the previous section, the first step was the creation of the list of key words to be used in the study. As already noted, since they are key words, most of them can be translated directly. In some cases, however, problems arise because the direct translation does not work well or because the translated term generates noise on making reference to words in radically different contexts. For these cases, we opted to follow one of the two alternatives below:

1. In those cases where, even if the direct translation is not valid, there is an equivalent word or expression, we treat this equivalence as valid.
2. When the direct translation is not valid and there is no equivalent word or expression in the same context, we eliminate that word from the list.

The list of words with their respective weights used in this study is compiled in Table 2. These key words are the ones that were used to recover the "tweets." According to the considerations in the previous section, in the capture of data the previously mentioned algorithm to generate the ranking was applied. Likewise, the correction coefficient based on Twitter use was applied; this information is shown in Table 3. This is how a ranking of feelings of happiness was obtained for Latin American countries according to Twitter data, as appears in Table 4.

TABLE 4. HAPPINESS RANKING ACCORDING TO TWITTER

City/Country	**Score**
Caracas (VENEZUELA)	582,461.34
Buenos Aires (ARGENTINA)	462,247.10
Bogota (COLOMBIA)	346,830.89
Mexico (MEXICO)	329,093.01
Santiago (CHILE)	244,561.10
Asunción (PARAGUAY)	179,175.01
Montevideo (URUGUAY)	83,808.53
Guatemala (GUATEMALA)	77,924.13
Lima (PERU)	62,399.62
Panama (PANAMA)	61,444.27
San Jose (COSTA RICA)	34,071.67
Santo Domingo (DOMINICAN REP.)	32,397.79
Quito (ECUADOR)	21,650.52
Tegucigalpa (HONDURAS)	14,668.87
Managua (NICARAGUA)	10,229.80
La Paz (BOLIVIA)	2,521.12

Source: compiled by the authors

If we compare these results to the ranking obtained on the basis of the surveys from Latinobarómetro (Table 1), we see that there are notable discrepancies, especially in some

countries with relatively high rates of Twitter use (Argentina, Chile, Paraguay, Uruguay and Peru; see Table 3). It seems as though in those countries where the use of Twitter is greater, there is a strong upward bias, such that they appear in relatively high positions in the happiness ranking presented in Table 4. This might be because social networks have a viral, disseminating effect, so both positive and negative messages are spread, and as a result, the values are much more extreme than the simple indicative proportion of number of users. This could also be interpreted to signify that countries with a greater number of users not only have more users, but also more active users. On the contrary, some countries with relatively low coefficients of Twitter use like Costa Rica, Panama and the Dominican Republic, precisely because of the absence of the aforementioned viral effect, occupy relatively low positions in the ranking shown in Table 4, while in the ranking made on the basis of Latinobarómetro (Table 1), they are in high positions.

As part of the experiment and with the aim of finding a correction factor that encourages making future evaluations at different temporary moments and including other factors, we decided to calculate a weighting or adjustment factor that would allow us to equate the results obtained through the use of social networks with those derived from the Latinobarómetro surveys. One justification for calculating this weighting factor is to try to offer additional information that contributes to explaining the differences between using both methods to infer the happiness of Latin Americans. In considering this weighting factor, it is observed that the weighting necessary to adjust the result is greater in smaller countries with lower rates of Internet and social network use, which supports the previously formulated hypothesis for explaining the differences between the ranking of Tables 1 and 4.

In analyzing the content of Table 5, the case of Bolivia deserves to be highlighted. Although its position in the ranking with data from social networks (18th in Table 4) is not very different from its position in the Latinobarómetro ranking (15th in Table 1), on a quantitative level, it presents a very big lag compared to the other countries. Everything seems to indicate that once again we see a polarization of the results owing to the scant use of social networks in this country.

TABLE 5

ORD1	ORD2	City/Country	Score	Factor	%Factor Obj.	Objective	Weight.Fact.	Final Score
11	1	San Jose (Costa Rica)		34,071.67	0.0023867	0.239%	81.32 17.68	602,370.3704
1	2	Caracas (Venezuela)		582,461.34	0.0001353	0.014%	78.81 1.00	583,777.7778
10	3	Panama (Panama)		61,444.27	0.0012481	0.125%	76.69 9.25	568,074.0741
3	4	Bogota (Colombia)		346,830.89	0.0002137	0.021%	74.13 1.58	549,111.1111
8	5	Guatemala (Guatemala)		77,924.13	0.0009493	0.095%	73.97 7.03	547,925.9259
4	6	Mexico (Mexico)		329,093.01	0.0002190	0.022%	72.07 1.62	533,851.8519
7	7	Montevideo (Uruguay)		83,808.53	0.0008536	0.085%	71.54 6.32	529,925.9259
12	8	Santo Domingo (Dom. Rep.)		32,397.79	0.0021650	0.216%	70.14 16.04	519,555.5556
15	9	Tegucigalpa (Honduras)		14,668.87	0.0046459	0.465%	68.15 34.41	504,814.8148
6	10	Asunción (Paraguay)		179,175.01	0.0003776	0.038%	67.65 2.80	501,111.1111
2	11	Buenos Aires (Argentina)		462,247.10	0.0001448	0.014%	66.94 1.07	495,851.8519
5	12	Santiago (Chile)		244,561.10	0.0002659	0.027%	65.03 1.97	481,703.7037
16	13	Managua (Nicaragua)		10,229.80	0.0059972	0.600%	61.35 44.42	454,444.4444
14	14	Quito (Ecuador)		21,650.52	0.0024549	0.245%	53.15 18.18	393,703.7037
18	15	La Paz (Bolivia)		2,521.12	0.0192890	1,929%	48.63 142.88	360,222.2222
9	16	Lima (Peru)		62,399.62	0.0006981	0,070%	43.56 5.17	322,666.6667

ORD1: Order based on twitter data.
ORD2: Order after applying the Weight Factor.

VI. CONCLUSIONS AND FUTURE RESEARCH

In this article, a first approach to measuring happiness in Latin America through the use of social networks is presented. Specifically, the social network used is Twitter, although we do not rule out the possibility of undetaking future studies with Facebook or other social networks. We have used Twitter because of its characteristics (ease of use, availability and popularity, geographical data, etc.). We have developed a process that permits the extraction of data and generation of a new ranking quickly and easily, which allows us to easily repeat the experiment with additional conditions, parameters and searches.

We can extract the following points as our main conclusions:

- The measurement of happiness through the use of social networks seems viable, and it is tremendously simple compared to traditional methods (e.g., surveys).
- The measurement of happiness through social networks like Twitter involves considering several factors in order to obtain reliable results. The most evident factors are the use of Internet and the use of social networks.
- The method used in this work consists of inferring the feelings of social network users on the basis of a semantic analysis of the words used in their communications and messages.
- It is possible to calculate, via objective and empirical means, factors that allow us to correctly interpret data collected through the use of social networks.

- In time, as the use of Internet and social networks increases, the use of these tools will be more precise.

As lines of future research, we propose the possibility of:

- doing new studies which incorporate data gathered over longer time periods
- including only countries with similar socio-economic conditions
- refining the creation of that weighting factor which could be converted into a rating
- including not only positive terms but also negative ones in order to improve reliability
- doing other studies that, instead of key words, are based on iconographic elements like "smiley faces."

REFERENCES

[1] Ahn, N. and Mochón, F. (2010). La felicidad de los españoles: Factores explicativos. Revista de Economía Aplicada, 54(XVIII), 5-31.

[2] Ahn, N., Mochón, F. and De Juan, R. (2012). La felicidad de los jóvenes. Papers Revista de Sociología. 97/2, 407-430.

[3] Argyle, M. (1999). Causes and correlates of happiness, in Kahneman, Diener and Schwarz (eds.): Well-Being: The foundations of hedonic psychology. New York: Russell Sage Foundation, op. cit., 353-373.

[4] Argyle, M. (2001). The psychology of happiness. New York: Routledge.

[5] De Juan, R and Mochón, F. La felicidad de los latinoamericanos en el periodo 2000-2009. (2012). Working paper 0212. Departamento de Análisis Económico. Uned.

[6] Di Tella, R., MacCulloch, R. (2006). Some Uses of Happiness Data in Economics. Journal of Economic Perspectives, 20(1): 25-46.

[7] Diener, E, Emmons, R.A, Larsem, R.J. and Griffin, S.A..The Satisfaction With Life Scale. Journal of Personality Assessment, 1985, 49, 1, pp 71-75.

[8] Dodds, P. and Danforth, C. (2009). Measuring the Happiness of Large-Scale Written Expression: Songs, Blogs, and Presidents. Journal of Happiness Studies. doi:10.1007/s10902-009-9150-9.

[9] Easterlin, R.A. (1995). Will raising the incomes of all increase the happiness of all? Journal of Economic Behavior and Organization, 27, 35-47.

[10] Easterlin, R.A. (2001). Income and Happiness: Towards a unified theory. The Economic Journal, 111, 465-84.

[11] Fowler, J.H, Christakis, N.A.Steptoe. and Roux,D. (2009) A Dynamic Spread of Happiness in a Large Social Network: Longitudinal Analysis of the Framingham Heart Study Social Network British Medical Journal. Vol. 338, No. 7685 , pp. 23-27.

[12] Frank M., Mitchell L., Dodds P., and Danforth C. 2013. Happiness and the Patterns of Life: A Study of Geolocated Tweets. Scientific Reports 2013. Vol. 3, No: 2625, doi:10.1038/srep02625

[13] Frey, B.S. and Stutzer, A. (2002a). What Can Economists Learn from Happiness Research? Journal of Economic Literature, XL(June), 402-435.

[14] Frey, B.S. and Stutzer, A. (2002b). Happiness and Economics. Princeton: Princeton University Press.

[15] Fuentes, N. and Rojas, M. (2001). Economic Theory and Subjective Well-Being: Mexico. Social Indicators Research, 53(3), 289-314.

[16] Gerstenblüth, M., Melgar, N. and Rossi, M. (2010). Ingreso y desigualdad: ¿Cómo afectan a la felicidad en América Latina? (Working paper, 09/10). Montevideo: DECON.

[17] Graham., C. and Pettinato, S. (2001). Happiness, markets, and democracy: Latin America in comparative perspective. Journal of Happiness Studies, 2, 237-268.

[18] Graham, C. and Pettinato, S. (2002). Happiness and Hardship: Opportunity and Insecurity in New Market Economies. Washington, DC: The Brookings Institution.

[19] Kahneman D, Krueger A, David A. Schkade, Schwarz, N, Stone, A.. (2004). A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method. Science 3 December. Vol. 306 no. 5702 pp. 1776-1780))

[20] Kahneman, D., Diener, E. and Schwarz, N. (1999). Foundations of hedonic psychology: scientific perspectives on enjoyment and suffering. New York: Russell Sage Foundation.

[21] Knibbs K.(2013). Can 10 million tweets measure happiness? New research says so. http://www.digitaltrends.com/social-media/can-10-million-tweets-measure-happiness/.))

[22] Killingsworth M. (2012). The future of happiness research. Harvard Business Review. January February. pp. 88-89))

[23] Krueger, A. B. and Schkade, D.A. (2007). The reliability of subjective well-being measures. (Working Paper, 13027). Cambridge: National Bureau of Economic Research.

[24] Larson, R. Csikszentmihalyi, M. and Graef, R. (1980). Mood variability and the psychosocial adjustment of adolescents. Journal of Youth and Adolescence. 9 (6). pp. 469-490.

[25] Larson, R. and Csikszentmihalyi, M.(1983). The experience sampling method. I. H. T. Rels (ed.), Naturalistic approaches to studying social interaction new directions for methodology of social and behavioral science. pp.41-56. San Francisco- Josey-Bas.

[26] Layard, R. (2005). Happiness: Lessons from a new science. New York: Penguin.

[27] Pena López, J.A. and Sánchez Santos, J.M. (2009). Economía y felicidad: Un análisis empírico de los determinantes del bienestar subjetivo de la población. Universidad de la Coruña, mimeo.

[28] Rojas, M. (2006) Life Satisfaction and Satisfaction in domains of Life: It is a simple relatioship? Journal of Happiness tudies, 7, 467-497,

[29] Rojas, M. (2007). Heterogeneity in the relationship between income and happiness: A conceptual-referent-theory explanation. Journal of Economic Psychology, 28, 1-14.

[30] Rojas, Mariano (2011) El Bienestar Subjetivo: Su Contribución a la Apreciación y la Consecución del Progreso y el Bienestar Humano". Realidad, Datos y Espacio: Revista Internacional de Estadística y Geografía, 2(1), 64-77.`

[31] Wilson, C, Boe, B, Sala, A, Puttaswamy, K. P.N. and Zhao, B.Y. (2009). User Interactions in Social Networks and their Implications. Proceedings of ACM SIGOPS/EuroSys European Conference on Computer Systems (EUROSYS 2009) Nuremberg, Germany, April.

[32] Wang. G, Konolige. T, Wilson. C, Wang. X, Zheng. H, and Zhao. B. Y. (2013). You are How You Click: Clickstream Analysis for Sybil Detection. Proceedings of Usenix Security. Washington DC, August., Software Agents, Modeling Software with BPM, DSL and MDA.

Francisco Mochón received a M.A and a PhD in economics from Indiana University and from Universidad Autónoma de Madrid. At present is full professor at UNED and belongs to the advisory board of the multinational digital content company, ZED. His research interests are the economics of happiness and the digital economy.

Oscar Sanjuán is a Lecturer at the Computer Science Department of the University of Carlos III of Madrid Spain. He is Ph.D. from the Pontifical University of Salamanca in Computer Engineering. His research interests include Object-Oriented technology, Web Engineering, Software Agents, Modeling Software with BPM, DSL and MDA.

Happiness and Human Relations: The Role of Materialistic Values. An ABM Illustration

Mariano Rojas[1,2], Ignacio Ibarra-López[2]

Facultad Latinoamericana de Ciencias Sociales – Sede México, Mexico
2 CIIE - Universidad Popular Autónoma del Estado de Puebla, Mexico

***Abstract* — This paper argues that a person's happiness must be understood as a phenomenon that emerges not only from her individual condition but also from her place in society. Understanding that a person is socially immersed implies giving a greater role to social interactions and social structure. The paper presents a simple model to take into consideration the role of human relations. An agent-based model (ABM) is used to illustrate the implementation of the model in understanding people's happiness.**

***Keywords* — Happiness, social interactions, agent-based models, relational values, materialistic values.**

I. Introduction

As John Donne's poem states, *No man is an island, Entire of itself, Every man is a piece of the continent, A part of the main*, the individualistic paradigm is incorrect to understand social phenomena. It is important to distinguish between individuals and persons. Individuals are always portrait out of context; they seem to be standing up nowhere. On the contrary, persons are socially immersed, they are in society; as Ortega y Gasset used to say, they are in their circumstance.

The study of happiness requires taking into consideration that it is a living experience that happens to persons and not to individuals. Happiness is experienced by persons who are in society and who are living in their circumstance. Thus, the understanding of happiness requires from incorporating a person's context, which implies for the need on incorporating how people interact with others.

It is well-known that a person's happiness emerges from her personal characteristics as well as from her society's characteristics: the social structure and the social networks that exist.

This paper wants to emphasize the role that social interactions play in generating happiness within different value contexts. In specific, the paper studies how materialistic values influence the way rational agents end up following to pursue happiness. However, rather than following an individualistic approach, this paper recognizes that human relations do play an important role in the generation of happiness; in consequence, it is necessary to incorporate people's interactions into a model to understanding happiness. A simple model is presented which assumes that happiness emerges from the consumption of both economic and relational goods; the model recognizes that it takes two –or more- agents to generate gratifying economic goods. Thus, people do interact in the generation of relational goods and their happiness does not depend on their isolated decisions but also on what their fellows do.

An agent-based model is constructed to study how people's procurement of happiness within a social-interaction context ends up generating solutions to the allocation of time between the working (the generation of income to buy economic goods) and relating (consuming relational goods.

The paper is structured as follows: Section II presents the state of the art, discussing the importance of human relations in explaining people's happiness, and showing that social interactions do emerge as a consequence of these human relations. Section III presents a model to understand a person's happiness as a consequence of her social structure and her social interactions. The model explains how rational people end up allocating their limited endowment of time between working and relating; the model also assumes that people are statically rational while they are dynamically bounded-rational. Section IV presents an illustration of the model by using an ABM model. A final comment is made in section V.

II. State Of The Art. Happiness And Human Relations

A. The importance of human relations

Happiness research has shown that human relations are crucial for people's well-being. There are many kinds of relationships which emerge from the social organization people live in.

The social-capital literature stresses the instrumental relevance of human relations. It states that by fostering trust among people economic transactions are promoted and markets expand; a process that raises people's income [1]. Human relations, however, are important by themselves; they do not only contribute to raising people's income, they do also contribute to wellbeing through many channels [2], [3], [4]. For example, the Self-Determination-Theory school states that the satisfaction of some psychological needs may be as

DOI: 10.9781/ijimai.2014.253

important as the satisfaction of material needs. [5] view relatedness as a basic psychological need and mention that people's wellbeing declines when it is not appropriately satisfied.

Rojas [6] has shown that human relations play an important role in explaining the relatively high levels of happiness in Mexico. [7] shows the importance of the family and friendship domains of life in explaining people's satisfaction with life. People are socially immersed and, in consequence, their relation to others is crucial. The importance of human relations is such that William James argued in his Principles of Psychology that the worst punishment for somebody is not physical torture but to go by life being completely unnoticed by everyone else. As a matter of fact, many economists have recognized the importance of human relations, pointing towards some of their benefits, such as: the correspondence of sentiments [8], nurturing [9], moral support [10], and so on.

Fowler and Christakis [11] have shown that happiness is contagious. They use longitudinal data from New England to study the dynamic spread of happiness in a large social network, finding out that happy people tend to spread happiness through their social network, even beyond their first-level network. As a matter of fact, according to their results, there is likelihood for a happy person to positively impact on the happiness of friends, of friends of friends, and even of friends of friends of friends. Of course, unhappy people do also tend to spread unhappiness. As a consequence of this phenomenon, happiness and unhappiness tend to show up in clusters within a social network rather than showing a random distribution. In other words, social networks show an arrangement where happy people tend to move together while unhappy people do also tend to move closer to each other.

B. Relational goods

Some economists have started using the term relational goods to refer to those human relations that directly contribute to people's well-being [12], [13], [14], [15], [16], [17], [2], [3], [4]. They are referring to gratifying relations that satisfy some psychological and economic needs, such as: competence and self-esteem, autonomy and sense of being appreciated, and relatedness

The nature of relational goods is such that their contribution to well-being substantially declines if they are traded in markets. In other words, a larger contribution to people's well-being is made by genuine -rather than commercialized- relationships. In consequence, it can be stated that relational goods have intrinsic value because they do contribute to people's well-being. However, because relational goods are not traded in markets their value does not show up in market prices. Any attempt to commercialize relational goods would automatically diminish their value.

The production of relational goods is time intensive because building genuine and strong positive human relations usually requires time. It is for this reason that the production of relational goods competes with the production of economic goods (working to generate income in order to buy economic goods) in the allocation of a limited endowment of time. Of course, some overlapping between the production of relational and economic goods may exists, such as when good human relations do emerge in the place of work.

Standard economic theory makes no consideration of relational goods as arguments in the utility function; it assumes that utility depends on economic goods alone. However, relational goods have proven to be relevant for well-being, as well as for human motivation. Income is an irrelevant proxy for access to relational goods because it is in the nature of these goods that they cannot be purchased. Income is also an irrelevant proxy for the production of relational goods because their production is time intensive; thus, having more income does not mean that people can enjoy more relational goods, in special if greater income is attained through more hours at work. When relational goods are taken into consideration it becomes clear that income is not the only relevant input for well-being, that it is not a good proxy for utility, and that it does not fully capture a person's contribution to society.

C. Many kinds of human relations

There are many kinds of human relations. For example, the family is an ancient institution which performs many roles [18] and where many human relations emerge and evolve: with spouse or partner, with parents, which children, and with other family members. Relationships within the family are characterized by their strength, solidarity, and support, and, in general, they are expected to make an important contribution to the well-being of all family members. The importance of the family for well-being has been pointed out by [19: 393] who states that "In the case of the labour market the distribution of resources is based on competition and individual performance. The welfare states' redistribution is focused at solidarity between citizens. In the case of the family the principle is reciprocity and an informal contract between family members concerning responsibilities for the welfare of family members. There is a contract between spouses, between parents and their children, between adults and their elderly parents, and between adults and further relatives." Of course, family relationships emerge from other important kinds of relationships, like dating, courtship, and engagement relations. Relationships within the family evolve in complex manners, depending on many factors such as congeniality, occupations, economic situation, job opportunities, studies, children moving abroad or getting married, and so on.

At the social level there are many kinds of relationships, from relations with colleagues at work to relations with neighbors in the community and classmates at school. Friendship constitutes a general term which emphasizes relations which in general is considered to be positive for people's well-being; it refers to a special kind of close and warm relationship where people care to each other and where people interact, spend time together, and share some common interests. There are also sporadic but positive relationships, like those that emerge in a stadium or when using the public-transport system. Not all human relations are positive to well-

being, the term 'enemy' refers to an extreme case of relation, from which well-being is not expected to emerge.

D.Social structure and social interactions.

Persons are socially immersed and they interact with each other within a social structure and given their own social skills and resources. Thus, human relations do not emerge out of nothingness, they emerge from a given specific social structure which may promote, deter, or modify human relations.

The literature on social structures is relatively old [20], [21], [22]. However, research on the relationship between social structure, human interactions, and happiness is relatively new [23]. Even though happiness is an experience of the person, its understanding requires a perspective that moves beyond individualistic characteristics to also incorporate those characteristics of the social structure the person is immersed in. Most researchers have focused on how some social-structure characteristics correlate to people's happiness. For example, [24] study the correlation of some characteristics such as work participation, income distribution, and sociocultural integration on happiness. Other studies focus on social interactions in a specific place, such as at work [25]. These studies are inherently static and do not capture the nature of human interactions that lead to people enjoying greater or lower happiness. [26: 117] states that "The demographic and social structure of the community/society provide the basis for interactions that lead to satisfactions, subjective well being and the quality of life"

Udy [21] approaches a social structure on the basis of the following components: the individual, the group, the physical arrangement, the system, and cultural norms, values, and beliefs. It is within this structure that human interactions (relational goods) as well as the production of economic goods emerge. Different social structures may lead to different human relations and may affect people's well-being.

To study the role that the social structure plays in human behavior [27] proposes an 'embeddedness approach' which recognizes that people's actions are embedded into social relations. This approach leads to the development of interactions-based models of individual behavior. Some of these models follow a rational approach while others bend towards bounded-rationality behavior [28], [29], [30].

Social interactions do imply that a person's well-being does not only depend on her actions but also on the actions other persons make. Thus, it is impossible to understand a person's situation without a closer look at the system from which interactions emerge and in which interactions are shaped [31].

Social-interaction models provide many advantages with respect to standard economic models; for example: they characterize the feedbacks that exist within persons in a population, they allow for considering different behavioral rules beyond rational behavior, and they can even incorporate heterogeneity across personas [32], [33], [34].

E. Agent Based Models

Agent-based models (ABM) study social behavior on the basis of computational agents which can be modeled as homogeneous or heterogeneous, and which can interact among them and with their surrounding environmental conditions [35].

ABM models are inspired on Complexity theory [36], and they are deeply rooted on General Systems theory. In consequence, ABM models deal with adaptable complex systems where heterogeneous or homogeneous agents interact on the basis of non-linear specifications [37]. Many characteristics are incorporated into adaptable complex systems, such as time-dependence, self-organization, difficulty in anticipating equilibriums and emergence of aggregate qualities which cannot be foresee from individual behavior [38], [39], [40].

There are some similitudes between ABM and games, such as: the existence of players (agents), players's moves described in terms of decisions and strategies, a set of behavioral rules, and a pay-off schedule. However, ABM introduce new relevant characteristics, such as: very large numbers of players, many dimensions for modeling heterogeneity across players, an idea of space (geography) which is relevant for people's actions, learning and evolutionary processes, non-optimizing behavioral rules, and clear specification for time..

III. Model Features

A simple theoretical model is developed to study how people interact within a social network and how people's happiness emerges out of these social interactions in s given social structure. The goal is to understand people's allocation of time between working and relating in a model where people act motivated by the procurement of greater happiness.

F. Model sketch.

1. Persons derive happiness (H) on the basis of two domains of life: economic and relational. The economic domain of life refers to the satisfaction which can be attained by purchasing economic goods; income (Y) is the relevant variable reflecting a person's purchasing power. Time is required to generate income. The relational domain refers to the satisfaction which can be attained by interacting with other persons such as spouse or partner, children, friends, colleagues, neighbors and so on. Attaining gratifying human relations does require allocating time to interact to people.
2. A Cobb-Douglas specification is used to model the relationship between satisfaction in the economic and relational domains and H. The parameters of the Cobb-Douglas specification reflect the relative importance of the economic and relational domains in generating happiness. The Cobb-Douglas specification does imply a given elasticity of substitution between the economic and the relational domains of life.

3. Persons have limited endowment of time which they must allocate between the production of relational goods (generating gratifying human relations) and the generation of income (which will allow them to enjoy economic goods). Because the endowment of time is limited, people do face a trade off in the allocation of time between relational and economic goods.
4. The generation of relational goods is not a matter of each person alone, since it is required for other persons to also allocate their time to generate relational goods. The simultaneous willingness of at least two persons is required to generate gratifying relational goods. Allocating time to relational goods may be a waste of resources if nobody else in the near social circle is willing to do it. On the other hand, this may be a highly-rewarding strategy if other people are also willing to do it.
5. Finding good partners and friends is not an easy task, and not everything is under control in this venture. Random effects may play an important role in this regard. A geographical space representing the degree of closeness between persons –in their willingness for social interaction- can be imagined. Due to random factors some persons may begin their trajectory being closed to each other –and in a better position to generate gratifying relational goods- while other persons may begin far away from others and, as a consequence, face a greater cost in generating gratifying relational goods.
6. Of course, people do also take actions to move closer to other persons, in special to those persons they seem to like or be attracted to. However, trying to build gratifying human relations does require allocating time to this activity, which implies an opportunity cost in terms of the time that could be allocated to generate economic goods. Thus, actions people take to move closer to others –in the geographical/relational space- do imply an 'investment' (sacrifice of present consumption of economic goods) with an uncertain return (the reward does also depend on what other people do)
7. People's decisions are motivated by their expected happiness; however, it would be presumptuous –and probably unrealistic- to assume people act rationally. Basic heuristics can be assumed, such as evaluating only a few options at a time.

G.A Simple Model.

Persons derive happiness (H) on the basis of consumption of two goods: economic goods (E) and relational goods (R):

$$H = f(E,R) \qquad (1)$$

A Cobb-Douglas specification is assumed to generate H.

$$H = E^{\alpha} R^{1-\alpha} \qquad 0 \le \alpha \le 1 \qquad (2)$$

The importance of each domain is given by the parameter α. This parameter reflects the culturally-dependent values in the community. A simple model assumes that α is similar for everybody (a homogeneous population) The parameter α can be considered as a parameter reflecting the importance of materialistic values; as α moves closer to 1 economic goods become more important -and relational goods become less important- in generating happiness. In consequence, the value of α becomes important in studying how happiness emerges in materialistic and relational societies.

Persons have a limited endowment of time (T) which can be distributed between the two domains at a given substitution rate (e.g.: working to generate income leads to more E, while having more time to relate with people leads to more R). In consequence, there is a time constraint given by T as well as production functions which transform the time allocated to generate income (T_Y) into E, as well as the time allocated to generate relations (T_R) into R.

$$T = T_Y + T_R \qquad T_Y, T_R \ge 0 \qquad (3)$$

The time allocated to generate income generates economic satisfaction, while time allocated to generate relational goods generates relational satisfaction. For simplicity, it is assumed that:

$E=T_Y$

$R=T_R$

Thus, from equations (2) and (3) we get:

$$H = (T - T_R)^{\alpha} (T_R)^{1-\alpha} \qquad 0 \le \alpha \le 1 \qquad (4)$$

T_R becomes the only control variable in the equation; a person must decide its value on the basis of maximizing *H*. It is assumed that the person is rational in choosing T_R; in other words, the person chooses the time allocated to relational goods in order to maximize her happiness. The parameter α is assumed to be exogenous and given by cultural factors.

H.A geographical-relational map.

The geographical space is conceived as a squared map with m*m cells. Distance in this map is conceived as a relational distance, people who are close to each other in this map can develop a good relationship; however, good relationships will not emerge with people who are located farther away in this geographical space.

A person seeded in a specific cell will have a neighborhood given by the cells directly surrounding her (a Von Neumann neighborhood is assumed). In an m*m-cells map most people will have a neighborhood with 8 surrounding cells; those people placed in the border will have a neighborhood with 5 surrounding cells, and those placed in the corners will have a 3-cells neighborhood

I. Introducing social interaction: The production of relational goods

People can decide how much time to allocate to the production of relational goods; however, the quality of these goods does not only depend on the time they allocate to human relations but also on the existence of other people who are also willing to share time with them. The transformation of relational goods into happiness depends on other people's decisions. This implies for happiness being contingent on

social interactions which are not completely under any person's control.

Thus, the happiness attained by allocating time to relational goods does also depend on how many people 'are around in the neighborhood'; the greater the number of people in 'the neighborhood' the greater the happiness that can be attained by allocating time to relational goods. Thus, the following modification to the model in equation (2) is introduced in order to capture this important characteristic in the generation of gratifying relational goods:

$$H = E^{\alpha/n+1} R^{1-\alpha/n+1} \quad (5)$$

Where n refers to the number of other agents in the neighborhood, $0 \leq n \leq 8$

Once the time constraint expressed in equation (3) is taken into consideration, equation (5) becomes:

$$H = (1-T_R)^{\alpha/n+1} (T_R)^{1-\alpha/n+1} \quad (6)$$

In words, equation (6) implies that the greater the number of neighbors (people who are close to and from which good relations emerge) the greater the happiness a person can get by allocating time to human relations. In this simple model this also implies that having neighbors (nearby persons to generate high-quality relational goods with) implies a decline in the relative marginal contribution of economic goods to happiness.

J. Point of departure

At time t=1, N agents are randomly seeded in the m*m-cells map. For each agent (Ai) seeded in a specific cell c in the m*m-cells map there is an initial condition where the agent must choose TR in order to maximize her happiness. The maximization procedure is based on equation (6), given the parameter α and the value of n. Notice that because agents are randomly seeded in the map then the specific value of n for agent Ai (ni) is also random within a range from 0 to 8. The agent acts in a rational way when choosing TR, it will be that value that that maximizes her happiness

K. Decision Rule

The model introduces a decision rule where at each period t each agent Ai acts looking to pursue greater happiness. However, this is not a rational behavior because the agent does not evaluate an unlimited set of options; as a matter of fact the agent will act by looking at only one option at a time rather than by looking at multiple options in a simultaneous way. This assumption reflects bounded rationality and constitutes a heuristic (thumb rule).

First, at time t, agent Ai evaluates her current situation, which depends on the value of the parameter α and the value of ni (how many neighbors the agent has); the agent maximizes her happiness by choosing the optimal TR.

Second, agent Ai randomly chooses an empty cell in the neighborhood and then evaluates what her happiness would be at that cell. This evaluation is done by the agent under the assumption that everything else would remain constant. This is: the agent assumes that other agents will not move when she moves to the selected cell. If happiness is greater in the selected cell then the agent moves, if not then the agent remains in the same cell. If all cells in the neighborhood are occupied then the agent does not move.

Notice that an agent may decide to move in procurement of greater happiness but may end up with lower happiness. This may happen because a person's decision is based on the assumption that everything else remains constant; in other words, the agent cannot foresee nor incorporate when taking her decision what the other agents will do. In Kahneman's terminology, greater expected utility does not imply greater experienced utility because happiness also depends on what others do and this is not contemplated by the agent. This heuristic implies for agents' actions to be motivated by procuring greater happiness in a bounded-rationality way (in a dynamic process), while maximizing happiness in a rational way (in a static process at time t).

It is assumed that all agents follow this decision rule at time t and this sets the conditions for the situation at time t+1. Some agents decide to move to another cell from t to t+1 in procurement of greater happiness; while other agents do not move due to already having a full neighborhood (all cells in the neighborhood are occupied) or because the selected cell does not imply greater happiness). If two or more agents decide to move to the same cell then the program randomly selects one of the agents to move while the others remain in their cell.

At time t+1 the whole process is repeated. At t+1 all agents will choose that level TR that maximizes their happiness, and they will then decide whether it is convenient to move or not. This creates the conditions for t+2. The process can go on for many periods.

IV. Illustration. The Impact Of Relational Values

L. Changes in the degree of materialism

The parameter α reflects the predominant values in the society; a value of α closer to 1 indicates a materialistic culture where economic goods have a greater importance in people's happiness, while a value of α closer to 0 reflects the predominance of a relational culture where human relations have a greater importance in people's happiness. The following illustration studies the impact of changes in materialistic values on people's allocation of time. It is studied how people allocate their time between working and relating as materialistic values (the value of the parameter α) change. Hence, the illustration studies whether people end up looking for happiness through the consumption of economic goods or through the consumption of relational goods in a society where social interactions take place through human relations.

M. Working vs. relating. The allocation of time

The main variables to keep track of are TR (the time allocated to relational goods) and TY (the time allocated to generating income in order to consume economic goods). The values for TR and TY emerge from an optimizing process (of equation (6)), given the value of the parameter α, and the value

of n. The value of n is agent-specific, it is determined in a random way at time t=1, and then (for time t>1) it is determined by the decisions taken simultaneously by the N agents which aim for greater happiness in a bounded-rationality way.

There are different stochastic processes playing in the model; first, the initial seeding of the N agents in the m*m-cells map; second, the selection on a specific cell to be evaluated by each agent; third, the selection of a specific agent in those cases where two or more agents decide to move to the same cell.

N. The initial situation. Point of departure

At t=1 the following values are assumed:

T = 16 (Time available to be allocated between relating and working)

T_R = 8 (Time initially allocated to relational goods)

T_Y = 8 (Time initially allocated to generate income)

N = 100 (number of agents)

m =33 (geographical/relational space of *m***m*-cells map)

Other relevant information:

A_i: agent i, i = 1, . . . 100.

n_i (number of neighbors to an agent i; n_i is in between 0 and 8, it is a random variable at t=1, and it is the result of all agents decisions for t>1)

t = 1, . . . , 200 (number of periods under consideration)

$$H_{ti} = (1 - T_{R\_ti})^{\alpha / n_{ti}+1} (T_{R\_ti})^{1-\alpha / n_{ti}+1}$$

(happiness function to be optimized at any time t by any agent i)

O. Results

Netlogo is used to run the model. The parameter α is gradually changed from 0.01 to 0.99 by increments of 0.01. Thus, 99 different scenarios for materialistic values are constructed. Each scenario is run 100 times for 200 periods. Averages for TR and TY are computed at t = 200 across the 100 runs for each value of the parameter α. These averages are denoted as: MTR_200(α) and MTY_200(α). Table I shows the values for these averages for different values of the parameter α. Figures 1 and 2 show the behavior of these averages as the parameter α increases, this is, as society becomes more materialistic (and less relational).

TABLE 1
ALLOCATION OF TIME BETWEEN WORKING AND RELATING FOR DIFFERENT DEGREES OF MATERIALISTIC VALUES

Degree of materialistic values (α)	Average time allocated to relating $MT_{R\_200}(\alpha)$	Average time allocated to working $MT_{Y\_200}(\alpha)$
0.01	15.97	0.03
0.10	15.74	0.26
0.20	15.48	0.52
0.30	15.22	0.78
0.40	14.95	1.05
0.50	14.66	1.34
0.51	14.63	1.37
0.52	13.56	2.44
0.53	13.42	2.58
0.54	11.49	4.51
0.55	13.39	2.61
0.56	7.05	8.95
0.57	6.89	9.11
0.58	6.73	9.27
0.59	6.57	9.43
0.60	6.41	9.59
0.70	4.81	11.19
0.80	3.21	12.79
0.90	1.61	14.39
0.99	0.17	15.83

The results shown in Table I and in Figures 1 and 2 are not surprising, but there are some interesting issues to remark:

First, as expected, people tend to work more and to relate less as they become more materialistic. As people tend to obtain more happiness from consuming economic goods rather than from relating to other people it is reasonable for them to spend more time working in order to have enough income to buy the economic goods; this implies that less time is available to produce relational goods.

Second, the relationship between materialistic values and hours allocated to working/relating is not lineal. There seems to be a threshold value for the parameter α (at about 0.55) that implies substantial changes in the decisions agents end up taking. In societies with strong relational values people do optimize spending a few hours working and a lot of time in human relations (generating relational goods). Beginning from a highly relational society (values of the parameter α close to 0), as the society becomes more materialistic there are only small changes in people's optimal decisions in the allocation of time. People tend to spend most of its available time relating and just a few hours working. However, when the value of the parameter α reaches 0.56 substantial changes do occur in this society; an abrupt reduction in the time allocated to human relations occurs and people start working much more hours. It seems that when relational goods are not highly regarded (high value of parameter α) then social interactions do not promote the emergence of relational goods and the whole society gets into a different path towards happiness; emphasizing consumption rather than human relations

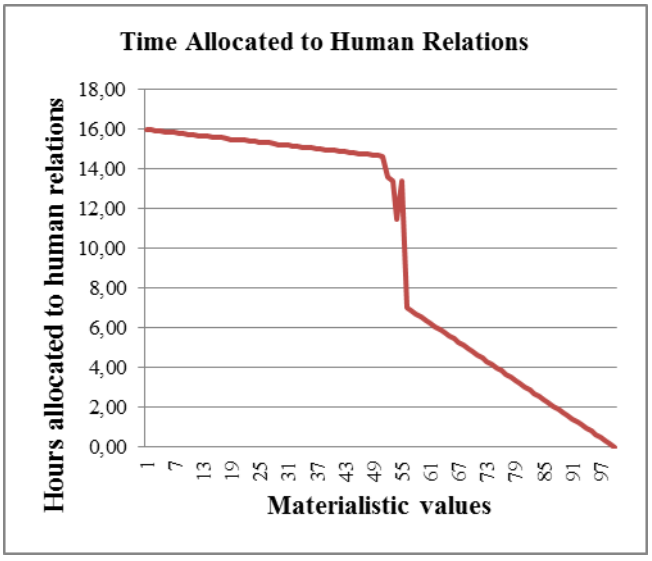


Fig. 1. Average time allocated to human relations as materialistic values become more important in a society. Average value for 100 runs of the time allocated to human relations after 200 periods.

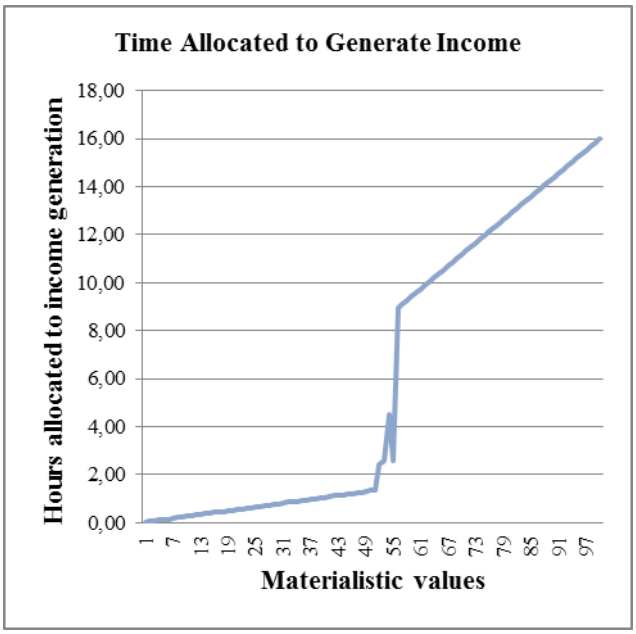


Fig. 2. Average time allocated to working (generating income) as materialistic values become more important in a society. Average value for 100 runs of the time allocated to human relations after 200 periods.

People's decisions on the allocation of time do also have important implications for the way progress is measured in societies. A materialistic society will end up producing a large quantity of economic goods but having little time to relate, while a relational society has a different way to attain happiness, where human relations are strengthened by social interactions and the society ends up with little production of material goods and a lot of time to enjoy gratifying human relations. It is clear that this behavioral strategy reflects in the indicators of production; under equal conditions, the Gross Domestic Product (GDP) tends to be higher in materialistic than in relational societies; however, it would be a big mistake to associate a greater GDP to greater happiness. In fact, a greater GDP is the result of a society becoming more materialistic, which leads to a different way of attaining happiness. In a similar way, it would be a big mistake to associate a lower GDP to lower happiness, since this may result from a society which finds optimal to attain happiness through the production of relational goods. As a matter of fact, what these findings suggest is that social progress should not be measured on the basis of GDP alone; relational goods should also be taken into account in order to understand people's happiness.

V. Final Comment

This paper has presented a simple model to illustrate the importance of social interactions in explaining people's happiness and in explaining how people pursue their happiness. Agent-based models can be used to understand complex situations where agents' decisions are contingent on other agents' actions. In these circumstances the general results for the society cannot be derived on the basis of studying individual behavior and do require a different perspective, where it is understood that persons are socially immersed.

Further research will sophisticate the model in order to get a richer understanding of how people's happiness emerges in a society.

References

[1] [1] J. Coleman, "Social Capital", in Foundations of Social Theory, J. Coleman, Comp., The Belknap Press of Harvard University Press, pp. 300-324, 1990.

[2] [2] L. Becchetti and M. Santoro "The wealth-unhappiness paradox: a relational goods/baumol disease explanation", in Handbook on the Economics of Happiness, L. Bruni, and P.L. Porta, Eds., Cheltenham, UK: Elgar, pp. 239-261, 2007.

[3] [3] M. Pugno, "The subjective well-being paradox: a suggested solution based on relational goods", in Handbook on the Economics of Happiness, L. Bruni, and P.L. Porta, Eds., Cheltenham, UK: Elgar, pp. 263-288, 2007.

[4] [4] B. Gui and R. Sugden, Eds., Economics and Social Interaction: Accounting for Interpersonal Relations, Cambridge, UK: Cambridge University Press, 2005.

[5] [5] E. Deci and R. Ryan, Intrinsic Motivation and Self-Determination in Human Behavior, New York: Plenum Press, 1985.

[6] [6] M. Rojas, "Happiness in Mexico: the importance of human relations", in Happiness across Cultures: Views of Happiness and Quality of Life in Non-Western Cultures, H. Selin and G. Davey, Eds., Springer, pp. 241-252, 2012.

[7] [7] M. Rojas, "The complexity of well-being: a life-satisfaction conception and a domains-of-life approach", in Researching Well-Being in Developing Countries: From Theory to Research, I. Gough and A. McGregor, Eds., Cambridge University Press, pp. 259-280, 2007.

[8] [8] A. Smith, The Theory of Moral Sentiments, Oxford University Press, 1759/1976.

[9] [9] C. Phelps, "The evolution of caring", in Economics & Happiness: Framing the Analysis, L. Bruni and P.L. Porta, Eds., Oxford University Press, pp. 267-302, 2005.

[10] [10] R. Sugden, "Correspondence of sentiments: an explanation of the pleasure of social interaction", in Economics & Happiness: Framing the Analysis, L. Bruni and P.L. Porta, Eds., Oxford University Press, pp. 91-115, 2005.

[11] [11] J.H. Fowler and N.A. Christakis, "Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the

Framingham Heart Study", British Medical Journal, 337 (a2338), pp. 1-9, 2008.
[12] [12] C.J. Uhlaner, "Relational goods and participation: incorporating sociability into a theory of rational action", Public Choice, 62, pp. 253-285, 1989.
[13] [13] B. Gui, "Beyond transactions: on the interpersonal dimension of economic reality", Annals of Public and Cooperative Economics, 71, pp. 139-169, 2000.
[14] [14] F.C. Wolff and L. Prouteau, "Relational goods and associational participation", Annals of Public & Cooperative Economics, 75(3), pp. 431-463, 2004.
[15] [15] R. Lane, The Market Experience, Cambridge University Press, 1991.
[16] [16] R. Lane, R., The Loss of Happiness in Market Economies, Yale University Press, 2000.
[17] [17] L. Bruni and L. Stanca, "Watching alone: relational goods, television and happiness", Journal of Economic Behavior and Organization, 65(3-4), pp. 506-528, 2008.
[18] [18] M. Rojas, "Communitarian versus individualistic arrangements in the family: what and whose income matters for happiness?", in Advancing Quality of Life in a Turbulent World, R.J. Estes, Ed., Springer Verlag, pp. 153-167, 2006.
[19] [19] J. Vogel, "The family", Social Indicators Research, 64, pp. 393-435, 2003.
[20] [20] R. Merton, Social Theory and Social Structure, The Free Press, NY, 1968.
[21] [21] S. Udy, "Social structure: social structural analysis", in International Encyclopedia of the Social Sciences 14, D.L. Sills, Ed., The Macmillan Company and the Free Press, pp. 489-495, 1968.
[22] [22] S.L Rytina, "Social structure", in Encyclopedia of Sociology Vol 4, E.F. Borgotta and R.J.V. Montgomery, Eds., Macmillan, pp. 2822–2828, 2000.
[23] [23] I. Gough, "Human well-being and social structures: relating the universal and the local", Global Social Policy, 4(3), pp. 289-311, 2004.
[24] [24] M. Haller and M. Hadler, "How social relations and structures can produce happiness and unhappiness: an international comparative analysis", Social Indicators Research, 75, pp. 169-216, 2006.
[25] [25] F. Tschan, N. Semmer and L. Inversin, "Work related and "private" social interactions at work", Social Indicators Research, 67, pp. 145-182, 2004.
[26] [26] A. Ferris, "A theory of social structure and the quality of life", Applied Research in Quality of Life, 1, pp. 117-123, 2006.
[27] [27] M. Granovetter, "Economic action and social structure: the problem of embeddedness", American Journal of Sociology, 91(3), pp. 481-510, 1985.
[28] [28] J. Coleman, Foundations of Social Theory, Harvard University Press, 1990.
[29] [29] C. Manski, "Economic analysis of social interactions", Journal of Economic Perspectives, 14(3), pp. 115-136, 2000.
[30] [31] W.A. Brock and S. Durlauf, "Discrete choice and social interactions", NBER Technical Working Paper No 258, 2000.
[31] [32] L.E. Blume and S. Durlauf, "The interactions based approach to socioeconomic behaviour", Working Paper No1, Social Systems Research Institute, University of Wisconsin, 2000.
[32] [33] L. Blume and S. Durlauf, "Identifying social interactions: a review", in Methods in Social Epidemiology, J. Oakes and J. Kaufman, Eds., Jossey-Bass, San Francisco, 2006.
[33] [34] L. Blume, W. Brock, S. Durlauf, and Y. Ioannides, "Identification of social interactions", in Handbook of Social Economics Vol. 1B, J. Benhabib, A. Bisin, and M. Jackson, Eds., Elsevier, pp. 853-964, 2011.
[34] [35] X. Hao Li, "Standardization for agent-based modelling in economics", Department of Economics and Social Sciences (DiSES), Polytechnic University of Marche, Ancona, Italy. MPRA Paper No. 47396, 2013.
[35] [36] G.W. Flake (1998) The Computational Beauty of Nature: Computer Explorations of Fractals, Chaos, Complex Systems, and Adaptation, MIT press, 1998.
[36] [37] L. An, "Modeling human decisions in coupled human and natural systems: review of agent-based models", Ecological Modelling, 229, pp. 25-36, 2012.
[37] [38] R. Axelrod and M.D. Cohen, Harnessing Complexity: Organizational Implications of a Scientific Frontier, Basic Books, 2000.
[38] [39] J.H. Holland, Emergence: From Chaos to Order, Reading, MA: Perseus Books, 1998.
[39] [40] S.M. Manson, "Simplifying complexity: a review of complexity theory", Geoforum, 32(3), pp. 405-414, 2001.

Mariano Rojas is professor of economics at the Facultad Latinoamericana de Ciencias Sociales – sede México and at Universidad Popular Autónoma del Estado de Puebla, México. He holds M.A. and PhD degrees in economics from The Ohio State University, United States. His main areas of research are happiness, quality of life, poverty, and progress of societies.

Ignacio Ibarra-López is associate professor of economics at Universidad Popular Autónoma del Estado de Puebla, México. He holds a PhD in economics from Universidad de las Américas, Puebla. His areas of research are complex systems, econometrics and political economy.

Graph-based Techniques for Topic Classification of Tweets in Spanish

Héctor Cordobés[1], Antonio Fernández Anta[1], Luis F. Chiroque[1], Fernando Pérez[2], Teófilo Redondo[3] and Agustín Santos[1]

1. IMDEA Networks Institute, Madrid, Spain
2. U-Tad, Madrid, Spain
3. Factory Holding Company 25, Madrid, Spain
4. Universidad Carlos III de Madrid, Spain

***Abstract* — Topic classification of texts is one of the most interesting challenges in Natural Language Processing (NLP). Topic classifiers commonly use a bag-of-words approach, in which the classifier uses (and is trained with) selected terms from the input texts. In this work we present techniques based on graph similarity to classify short texts by topic. In our classifier we build graphs from the input texts, and then use properties of these graphs to classify them. We have tested the resulting algorithm by classifying Twitter messages in Spanish among a predefined set of topics, achieving more than 70% accuracy.**

***Keywords* — Topic classification, text classification, graphs, natural language processing**

I. INTRODUCTION

Topic classification of texts is one of the most interesting challenges in Natural Language Processing (NLP). In the field of the happiness research it is important to combine sentiment analysis with topic classification techniques, in order to determine the reasons why a subject expresses happiness or sadness. The problem is to assign to every input text to be classified one topic chosen from a collection of predefined topics. Topic classifiers have commonly used a bag-of-words approach, in which the classifier uses (and is trained with) selected terms from the input texts. In these types of approaches the biggest issue is that the set of potential terms used is huge, and has to be reduced to have a practical classifier. Hence, the preprocessing of the texts and the selection of the most important terms to be used becomes fundamental.

In this work, we present classification techniques that are not based on the bag-of-words paradigm. Instead, they generates graphs from the texts, and use graph similarity to classify them by topic. The resulting classifier uses much fewer attributes than bag-of-words classical classifiers.

A prototype classifier was developed using the techniques proposed here, and was used to participate in the topic classification challenge of the Workshop on Sentiment Analysis at SEPLN - 2013, known as TASS 2013 (*Taller de Análisis de Sentimientos en SEPLN 2013*). As in previous years, the challenge organizers prepared and made available a data set for evaluation. For topic classification, a set of Twitter messages (tweets) in Spanish were provided. Some of these tweets had been previously classified among predefined categories (politics, economy, music, sports, etc.), and the rest was to be classified by the systems developed by the challenge participants. The classifier we developed ended in 3rd position (with respect to the F1 characteristic), very close to the systems that ended first and second, which used classical techniques.

Additionally, we have also tested different configurations of our classifier using the whole data set of tweets provided by the TASS organizers (including the ones used for evaluation), and found that our classifier achieves accuracies above 70%, using very few attributes. In the classifier developed and tested in this work, we have also explored pre-processing alternatives, such as simple Named-Entity Recognition, Thesauri and specific dictionaries (e.g., SMS abbreviations) to account for the special medium Twitter is. We believe that thorough work on this pre-made knowledge data bases could greatly improve the results of the classification.

The rest of the paper is structured as follows. We revise graph-based approaches for NLP in Section II. In Section III we describe the basic techniques used by our classifier, while in Section IV we describe how these techniques have been transformed into an operational system. In Section V we present the evaluation results that have been obtained and discuss their significance and implications.

II. STATE OF THE ART

The great representational power of graphs, in terms of element relationships, and the extensive mathematical work in graph theory, have been useful for text processing. Graph techniques have been successfully exploited for many tasks such as text summarization and information retrieval.

In fact, a number of scientific works use graph techniques for text summarization of big documents, such as [2] or [16]. Similarly, the TextRank method [10], which is the application

Partially funded by the SOCAM research project, Spanish Ministry of Industry, Energy and Tourism.

DOI: 10.9781/ijimai.2014.254

of the well-known PageRank metric [3] to text graphs, has been used with remarkable success [7] to extract good representatives in text-related graphs by using a random-walk approach. The method is based on the assumption that well-connected nodes (e.g., terms or sentences), would be good representatives of a graph. These works also use an additional set of techniques in order to exploit the relation between sentences in the same document. For this matter, methods such as tf-idf [14], combined with mutual information, information gain, Helmholtz principle [4], and other weighting mechanisms, have been developed to fine-tune the importance of the terms, mainly towards a subsequent bag-of-words scheme. For example, for classification tasks, it is common to describe documents within a Vector Space Model (VSM), and classify them with Rocchio or SMO classifiers, in which each feature is a weighted term. These methods relay in calculating centroid representatives of the text to summarize. Unfortunately, they may sometimes fall in a multi-centroid problem, for which good decision borders determination can be difficult to solve.

In this work, we propose a system where very short text classification is possible by using a vector classification model for which the features are not terms, but graph metrics, thus significantly reducing the training and exploitation computational requirements, while retaining reasonable accuracy. As mentioned, this work makes use of the TASS2013 corpus, managed by SEPLN (Spanish Society for Natural Language Processing) for its TASS sentiment classification challenge. This corpus is in Spanish, which prevents us from using well-known baselines for the English language, such as Reuters-21578 [9]. Instead, we will compare ourselves with other participants in the same task.

Nevertheless, this work is a first step in the application of graph techniques to topic classification of short texts, so it must be taken as a proof of concept. More advanced techniques can be used in conjunction with this classification scheme, such as PoS tagging and dependency trees [17], or sophisticated text normalization [13].

III. Basic Graph-based Classification Techniques

The basic principle for all our techniques is that every piece of text (tweets in this case, and in general a sentence) can be represented as a graph. Essentially, for a given text our proposal uses the words in the text as graph vertices (we usually work only with the word lemmas, and optionally with named entities), and creates weighted edges between the words. We have considered different ways of assessing weights on the edges. A simple option is that the weight represents the frequency with which both words occur together in the text. Another more sophisticated (and complex) choice is that this frequency is weighted by the distance between the words in the syntactic tree of the text. There are other alternatives for building the graph that we deem of great interest in future work (especially those based in directed graphs).

Knowing how to build a graph for each tweet, the first hypothesis for our system is that graphs belonging to the same topic have a common representative structure (topic reference graph). For the text classification, we look for the similarities between the graph generated for a given text and different topic reference graphs. Hence, our work uses a technique of graph similarity in order to detect the topic of a piece of text.

Hence, for our experiments, we have built a reference graph for each topic. This graph is the union of all the graphs generated from all the texts of the same topic. In the resulting reference graph, the weights of the same edge in different graphs are added. This decision is based on the second hypothesis of our work, that is, all words relate to each other with different intensities depending on the topic. For instance, when the topic is Politics, the words *Presidencia* and *Congreso* will show a strong relationship. These same words may not appear or have a weak relationship in other topics (e.g., Football). Therefore, the reference (union) graph created for every topic is expected to be very different. The overall process of building the reference graphs is shown in fig. 1.

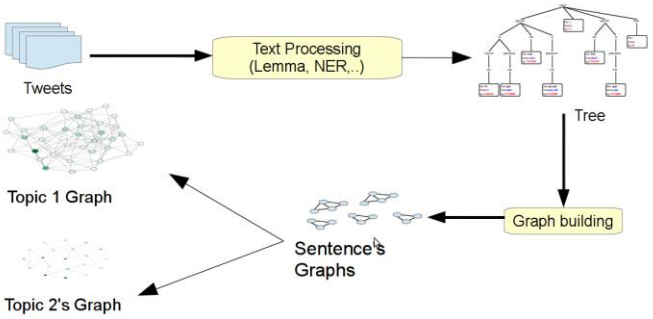


Fig. 1: Graph building process

Hence, using a pre-classified set of tweets for training, our system builds the reference graph for each of the different topics. When a new tweet needs to be classified, its graph is generated. Then, we search for the reference graph with the highest similarity with the tweet graph we want to classify. Fig. 2 shows this process.

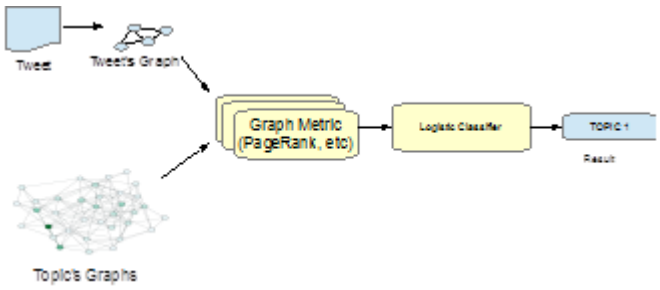


Fig. 2: Tweet classification process

The basic mechanism previously described opens up a wide spectrum of choices and approaches that can be combined in multiple ways. The first step in the mechanism is to build the graph for the tweet. As we have already mentioned, in our

work we have explored several options for selecting nodes and assigning weights to the edges. Similarly, we have used several criteria to measure the similarity of a given graph to a reference graph. In the following sections we go into greater detail about the methods we have employed.

IV. Implementing the Classifier

In this section we describe how the classifier has been developed, and particularly how the techniques described in the previous section have been implemented. In Section 4.1 we describe the preprocessing that all the tweets go through before using them to build the associated graphs. Section 4.2 describes how the reference graphs get built. Finally, in Section 4.3 we describe how the topic of a new tweet is identified.

A. Preprocessing of the Text

As a step prior to building and analysing the graphs, we run a preprocessing phase on the texts. This is a typical step in many natural language processing techniques. In this phase, the text is corrected, analysed, and separated into simple elements. In our work we have used the Hunspell dictionaries to obtain an orthographically correct text. We have also used the dictionary of SMS abbreviations and symbols (SMS dictionary) that we already used in the system we developed for TASS 2012 [5]. In addition, we have used Freeling [12] for word lemmatization, taking always into account the automatic disambiguation of lemmas according to the syntactic function. Freeling is also used to parse the syntactic tree of the tweets, which is used for calculating the distances between words. These distances will be used in the following sections.

Another step in the preprocessing phase has been identifying the Named Entities (Named Entity Recognition or NER process). The objective in this step has been to have mechanisms available in order to unify in a single term collections of words that refer to the same concept (e.g., *Real Madrid*, *Real Madrid C.F.*). To this end, and as a proof of concept, we have used a small manually-created catalog of slightly less than 100 entity names, with several variations for each one. For the creation of the catalog, the texts in the training set have been separated into n-grams, with no limit as to their length, using the technique described in [11]. After the extraction of statistically significant n-grams, the catalog was manually extended both in similar concepts (for instance, the name of a media provider) and in the different ways these concepts may be present.

For the NER we have used a search in the catalog for every single occurrence of the n-gram in the text in order to verify if it refers to one of the entities in the catalog. If so, the n-gram gets substituted by a given canonical name. For instance, the bigram *Mariano Rajoy* has been considered as one such entity, in this case with canonical name *mariano_rajoy_brey*. The whole process has been executed as an experiment, and we believe that broadening its use and having a more complete catalog could improve significantly the quality of the results.

In summary, the preprocessing of each tweet goes through the following phases: first, all URL's are deleted from the tweet; second, using the SMS dictionary, the abbreviations and symbols present are replaced by their textual equivalent; third, orthography is corrected using the Hunspell dictionaries; fourth, the tweet language is detected using Cybozy Labs Language Detection Library [15] and, if it is not Spanish, it is discarded; fifth, NER is applied, substituting the entities found for their canonical name (this phase can be removed at will to check how effective it is in the overall result); sixth, lemmatization is performed using Freeling; seventh and last, all the stop words are removed.

B. Reference Graphs

The key process to build a reference graph per topic is the process of building a graph for each text, since the reference graph is the union of these graphs. We have tried several options to build text graphs, described below, some of them very involved. The differences are on the set of nodes included in the graph or the way weights are assigned to the edges of the graph.

The simplest option considered for building text graphs has been using as nodes of the graph the words of the text (or the named entities, if used). Then, two nodes are connected with an edge whose weight is the product of their respective number of appearances in the text. (For instance, if in a text the word *concierto* appears twice and the word *guitarra* appears three times, the nodes of these two words are connected with a link of weight 6.) The reference graph obtained with this option has as nodes all the words that appear in the tweets of the topic, and the weight of a link between two words is the number of instances of both words occurring together in the same tweet.

A second option explored assigns to the link between two words a weight that is inversely proportional to the distance between the two words in the text. The intuition is that two words occurring together in a text have larger affinity, and hence should have a stronger link, than words occurring at opposite ends in a sentence. This distance is derived from the syntactic parsed tree as produced by Freeling. To calculate the distance between two words we count the number of jumps in the parsed tree from one word to the other. Our experiments revealed that the results obtained with this option are similar to those with the previous one. Hence, this option was discarded, due to the additional complexity.

Another option that has been explored is using as node set not only the words that appear in the text but also all its synonyms provided by a thesaurus. The intuition is that this will increase the information of the resulting graph. In order to introduce a difference, the weight of the links involving synonyms was slightly below one, while the links connecting words in the text had weight one. In the tests run, the use of synonyms decreased the quality of the results, possibly because they interfered with the use of centrality measures for graph topic. We also tested the use of synonyms when trying to benefit from the graph information (not at the time of creating it). In this case we did not detect any significant

improvement either.

As mentioned, none of the options explored was sensibly better that the first option, which is also the simplest one. Hence, this is the type of text graph that is considered in the rest of the paper. However, we think that the use of weights based on distance and synonyms must be addressed in future work since we expect that the augmented graph obtained can improve the reference graphs, and consequently yield a higher rate of successful classifications. In fact, other works such as [1] have already benefited from using thesaurus information.

C. Text Classification

We describe now how the classification of an input text has been done. One of the main questions in our approach is related with the problem of detecting graph similarity. Electing the measure of similarity is a complex decision since there are a great variety of measures and it is not clear which one would be the most appropriate for our problem. In our work we have used several measures, but all of them use the subgraphs of the reference graphs obtained after filtering out the words that do not occur in the text to be classified. That is, for each reference graph we have extracted the words occurring in the text, and we keep the links between them (i.e., we obtain the subgraph induced by the words of the text). Thus, for each topic we obtain a topic subgraph that can even be empty if no word in the text is found within the reference graph.

The following step is to determine one or several topology measures that, when applied to the topic subgraphs, would allow us to choose the topic(s) of the text. We have used two large types of measures: those based in node metrics and those based in relations metrics. The node metrics have mainly been just two: PageRank [3] and HITS [8]. For the computation of these metrics we have used the variants for undirected graphs with weighted links, and applied them to the topic reference graphs. As a result, each node of the reference graph is assigned a measure (its PageRank or HITS values).

Unfortunately, the size of the reference graphs is heavily influenced (biased) by the training set (i.e., number of tweets for each topic), and the centrality measure assigned to the nodes are influenced by the size of the graph. Hence, we attempt to compensate this deviation by means of a normalization of the centrality measures. Following a simple hypothesis, we assume that, given equal representation, the values for the centrality measures would decrease according to the number of graph nodes. Hence, we have normalized the number depending on the size of the reference graph of a given topic. On the other hand, since these values are also dependent on the graph topology in an unpredictable way, we have tried using non-linear operations (particularly, powers like 0.5 or 1/3), in order to give more representation capability to the system.

Then, once the topic subgraph has been extracted for a text, the topic is assigned a value that is the sum of the measures of the nodes of the subgraph (for instance, the normalized sum of PageRank for all the nodes in the subgraph). Computing this value is fast and simple from the precomputed reference graphs. These centrality measures (PageRank and HITS) have been very useful in determining the text topic, as we show later in Section V (see Table I). We observed no big differences between using PageRank and HITS.

As a first approach the value assigned to each topic could be directly used for classification. After adding up the centrality measures for each word in a topic subgraph, the text is classified to the topic with the highest value. With this methodology we achieve nearly a 60% of correct classifications. However, using more sophisticated classifiers (provided in Weka) we achieve a higher rate of accuracy, as we show below.

In addition to the centrality measures, our work has also contemplated links measures. Since every link has a weight, we can compute metrics using those values. We have tried several techniques, but all of them are based on the density of the topic subgraphs (a weighted sum of the links weights). This technique by itself has not rendered better results, but during the evaluation with the training set the technique has proved to be fundamental when combined with the other techniques described before.

In order to combine all the measures described, we have used classifiers included in the Weka system [6]. Each tweet was represented by a vector formed by all the available metrics (PageRank's sum, HITS' sum, graph density, etc.) for every topic reference graph. All in all we have a vector with up to 70 numeric values at our disposal. Of all the classification methods available in Weka, we found that the family of Logistic produced a higher rate of correct classifications. Especially the Logistic MultiClass Classifier method, appeared to give better results in a consistent way over the training set. Hence, all the results shown below use this classifier.

V. Results and Discussion

We have evaluated our system with different configurations. In all the runs we have trained Weka with the full training set of TASS 2013 (approximately 7,000 tweets) and we have assessed the resulting model with slightly less than the 60,000 tweets of the test set (leaving out some tweets we could not obtain). Weka's algorithm in use has always been SimpleLogistic, as mentioned above.

In Table I we show the results in all the runs. The column "Configuration" shows the text attributes used: PageRank (PR), HITS, graph density (GD), and the modifications applied. These attributes have been generated for every single tweet both during training and evaluation. The column NER shows whether entity recognition has been used or not. As mentioned before, we have disabled this feature in some runs to measure the variation in results. The column "Accuracy" shows in percentage how the system identifies a tweet as belonging to one given topic, according to the evaluation data supplied. Experiment 1 shows the configuration submitted to the TASS 2013 contest.

Tables II and III show information about the distribution of the categories, both for the entry tweets and the results of the

classifier used in experiment 1. Note that some tweets belong to more than one category, so for the sake of clarity we have expressed both the occurrence rate, and a normalized occurrence rate. This latter is intended to express the occurrence rate as though the sum of occurrences was 100%.

We present the results by category instead of showing a confusion matrix, because the possibility of finding several categories for one tweet would make the latter large and unintuitive. In Table III the success rate must be interpreted as the proportion of the tagged predictions within the category whose tweet belongs, at least, to that category.

TABLE I

Experiment	Configuration	NER	Accuracy (%)
1	$PR^{0.5}$, PR, PR^2, $HITS^{0.5}$, HITS, HITS, GD	Yes	71.90
2	$PR^{0.5}$, PR, PR^2, $HITS^{0.5}$, HITS, $HITS^2$	Yes	71.62
3	$PR^{0.5}$, PR, PR^2, $HITS^{0.5}$, HITS, $HITS^2$	No	71.38
4	PR	Yes	69.78
5	$PR^{0.5}$	No	69.45
6	$PR^{0.5}$	Yes	71.64
7	$PR^{1/3}$	Yes	71.58
8	$PR^{0.1}$	Yes	69.04
9	HITS	Yes	69.75
10	$HITS^{0.5}$	Yes	71.32
11	$HITS^{1/3}$	Yes	71.35
12	$HITS^{0.1}$	Yes	68.88

TABLE II

Topic	Tweets	Occurrence (%)	Normalized occurrence (%)
movies	596	1.0	0.9
sports	135	0.2	0.2
economy	2549	4.2	3.7
entertainment	5421	8.9	7.8
football	823	1.4	1.2
literature	93	0.2	0.2
music	1498	2.5	2.1
other	28191	46.4	40.5
politics	30067	49.5	43.2
technology	287	0.5	0.4

From the results presented we think that the centrality metric used (PageRank or HITS) does not incur significant difference. On the contrary, the use of a specific normalization may represent a significant improvement (around 2%, for instance, between Experiments 4 and 6). This, together with the good results achieved by using centrality metrics, leads us to believe that choosing an appropriate normalization is of paramount importance for the improvement of results, or in any case, using a metric capable of taking all the factors (size, topology, etc.) into account. We believe that this is an interesting area for future research.

During the execution of the experiments we have detected sensitivity to the available vocabulary. Topics with very few tweets tended to be ignored, such as the case of Technology, because the generated reference graphs are not representative enough. One possible future work could focus on evaluating the sensitivity with larger training sets, and thus determining and measuring how important this effect may be.

TABLE III

Topic	Predictions	Ratio vs. total	Accuracy rate
movies	460	0.77	43.26
sports	67	0.11	47.76
economy	612	1.03	50.16
entertainment	6919	11.66	38.98
football	420	0.71	52.62
literature	60	0.10	25.00
music	1095	1.84	51.60
other	19753	33.29	77.00
politics	29890	50.38	78.27
technology	58	0.10	32.76

In a similar way, this sensitivity could be tested enlarging the NER collection dictionary, so that it can represent in greater detail the topics that the system handles. Maybe given the very limited size of the dictionary used (less than 100 entity names), the impact in the results is not very significant, although consistent (around 0.3%). We should also consider that the NER rate is about 18.3% (occurring rate per tweet) and, as the corpus tweets have been selected, not many different NE recognitions have occurred. Thus we may hypothesize that the impact could be greater within more heterogeneous corpora and bigger dictionaries. This topic is worth to be explored further.

Additionally, the use of the Graph Density in Experiment 1 combines well and was able to improve another 0.3% over the already complex combination of PageRank and HITS in Experiment 2. Nevertheless, it has to be noted that it is not worth increasing unnecessarily the number of characteristics, because as it is shown on Experiment 6, some well chosen metric may be very significant by itself.

The automatic evaluation of the predictive models in Weka is limited because it cannot take more than one prediction per vector, whereas the tweet labelling may include more than one topic per tweet. It is quite possible that a classifier that allows more than one topic per tweet would achieve better results.

Concerning the results for individual categories (Table III), the system appears quite biased towards the main categories (politics and other), as they account for 46.4% and 49.5% respectively of the original tweets. In these cases the system achieves roughly a 78% of correct classification. However, the remaining categories show a rather poor behaviour, many below a mere 50%. Of particular note is the case of entertainment, with a success rate of only 38%, even though it is the third category in the total number of tweets.

We think that an additional experiment with more accurate training could reveal if this behaviour is due to an unbalanced training or to the actual design of the system. Since the number of training texts in some categories (for instance, literature) is rather scarce, we think that a far more complete training set than that currently available would be needed.

REFERENCES

[1] Aseervatham, Sujeevan. 2007. Apprentissage à base de Noyaux Sémantiques pour le traitement de données textuelles. Ph.D. thesis, Université Paris-Nord-Paris XIII.

[2] Blanco, Roi and Christina Lioma. 2012. Graph-based term weighting for information retrieval. Information retrieval, 15(1):54-92.

[3] Brin, Sergey and Lawrence Page. 1998. The anatomy of a large-scale hypertextual web search engine. Comput. Netw. ISDN Syst., 30(1-7):107-117, April.

[4] Dadachev, Boris, Alexander Balinsky, Helen Balinsky, and Steven Simske. 2012. On the helmholtz principle for data mining. In Emerging Security Technologies (EST), 2012 Third International Conference on, pages 99-102. IEEE.

[5] Fernández Anta, Antonio, Luis Núñez Chiroque, Philippe Morere, and Agustín Santos. 2013. Sentiment analysis and topic detection of Spanish tweets: A comparative study of of NLP techniques. Procesamiento del Lenguaje Natural, 50:45-52.

[6] Hall, Mark, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. 2009. The WEKA data mining software: an update. SIGKDD Explorations, 11(1):10-18.

[7] Hassan, Samer, Rada Mihalcea, and Carmen Banea. 2007. Random walk term weighting for improved text classification. International Journal of Semantic Computing, 1(04):421-439.

[8] Kleinberg, Jon M. 1999. Authoritative sources in a hyperlinked environment. J. ACM, 46(5):604-632, September.

[9] Lewis, David D. 1997. Reuters-21578 text categorization test collection.

[10] Mihalcea, R. and P. Tarau. 2004. TextRank: Bringing order into texts. In Proceedings of EMNLP-04and the 2004 Conference on Empirical Methods in Natural Language Processing, July.

[11] Nagao, Makoto and Shinsuke Mori. 1994. A new method of n-gram statistics for large number of n and automatic extraction of words and phrases from large text data of Japanese. In Proceedings of the 15th conference on Computational Linguistics, COLING 1994, Volume 1, pages 611-615. Association for Computational Linguistics.

[12] Padró, Lluís, Samuel Reese, Eneko Agirre, and Aitor Soroa. 2010. Semantic services in freeling 2.1: Wordnet and ukb. In Principles, Construction, and Application of Multilingual Wordnets, pages 99-105, Pushpak Bhattacharyya, Christiane Fellbaum, and Piek Vossen, editors, Mumbai, India, February. Global Wordnet Conference 2010, Narosa Publishing House.

[13] Porta, Jordi and José Luis Sancho. 2013. Word normalization in twitter using finite-state transducers. Proc. of the Tweet Normalization Workshop at SEPLN 2013. IV Congreso Espa nol de Informática.

[14] Salton, Gerard and Michael J McGill. 1983. Introduction to moderm information retrieval.

[15] Shuyo, Nakatani. 2010. Language detection library for java. http://code.google.com/p/language-detection/.

[16] Thakkar, Khushboo S, Rajiv V Dharaskar, and MB Chandak. 2010. Graph-based algorithms for text summarization. In Emerging Trends in Engineering and Technology (ICETET), 2010 3rd International Conference on, pages 516-519. IEEE.

[17] Vilares, David, Miguel A. Alonso, and Carlos Gómez-Rodríguez. 2013. Una aproximación supervisada para la minería de opiniones sobre tuits en español en base a conocimiento lingüístico. Procesamiento del Lenguaje Natural, 51:127-134.

Héctor Cordobés obtained his MSc in Telecommunications in 2003 from Universidad Carlos III. From that year he worked for Motorola, participating in international projects such as Motorola Soft Switch IMS solution, High-Availability Computing Platform or Kreatel IPTV, as Systems Engineer and Developer. In 2013 he joined IMDEA Networks as a Research Engineer and he is currently studying a Research Master in Multimedia and Communications in Universidad Carlos III.

Antonio Fernández Anta (M'98–SM'02) is a Research Professor at Institute IMDEA Networks. Previously he was a Full Professor at the Universidad Rey Juan Carlos (URJC) in Madrid, where he has been on the Faculty since 1998. He was before on the Faculty of the Universidad Politécnica de Madrid, where we received an award for his research productivity. He has been a postdoc at the Massachusetts Institute of Technology from 1995 to 1997. He has almost 20 years of research experience, with a steady productivity of more than 5 papers per year on average. He has published in top conferences and journals like INFOCOM, STOC, FOCS, PODC, DISC, Journal of the ACM, SIAM Journal on Computing, or IEEE Transactions on Computers. He is vice chair of the Steering Committee of DISC and has served in the TPC of numerous conferences and workshops. He is a senior member of the IEEE since 2002 and of the ACM since 2007. Antonio Fernández Anta received his M.Sc. and Ph.D. degrees in Computer Science from the University of Louisiana in 1992 and 1994, respectively. He completed his undergraduate studies (Licenciado and Diplomado en Informática) at the Universidad Politécnica de Madrid, Spain, in 1988 and 1991 respectively, having received awards at the university and national level for his academic performance.

Luis F. Chiroque obtained his B.Sc. degree in Telematics Engineering from the Polytechnic University of Madrid. Currently, he is studying a Master's Degree in Mathematical Engineering at University Carlos III of Madrid and he is a PhD student at IMDEA Networks, where he previously worked in the SOCAM project. His main research interest areas are graph theory, social networks and big data.

Teófilo Redondo (BArts -1985, MArts - 1986; Universidad Complutense de Madrid - UCM) is Project Portfolio Coordinator at Factory Holding Copany 25. He was before Technology Architect & Director of Innovation Projectsat Universidad Internacional de La Rioja (UNIR). Previously he developed a career at IBM covering several areas like Cloud Computing and Big Data Architectures, Enterprise Solutions Architect (SAP, Oracle Solutions, Dassault Systemes), and as SOA Architect. He started in the Research Division of IBM with several projects on Machine Translation, during which he produced a number of articles on this subject. He was Visiting Scholar at Stanford University (1987). He currently holds an Associate Professorship at UNIR teaching about eLearning in Social Networks as well as Natural Language Processing techniques. He is affiliated

with SEPLN (Sociedad Española para el Procesamiento del Lenguaje Natural) since almost the beginning.

Fernando Pérez García, is a R&D project technical director at U-tad (Universidad de Tecnología y Arte Digital). He was before at IBM, as a Data Specialist, in the Business Analytics & Optimization service line, developing Business Intelligence and Data Mining projects for main spanish banks and carriers. Fernando is an entrepreneur, Android and Graphs enthusiast. Creator of Nautka Apps, Fernando has developed applications with over 100K users, providing natural language processing technologies to final users. Fernando is the inventor of the US Patent 7289025: "Method and system for securing an electronic device", on October 2007. Fernando completed his undergraduate studies (Licenciado en Informática) at the Universidad Pontificia de Salamanca in 1997.

Agustín Santos, PhD in Computer Science from Universidad Rey Juan Carlos de Madrid. He currently works at IMDEA Networks in several R+D projects. He previously developed his professional activities in the private sector. His broad experience covers technological base companies for which he has been founding partner. He has also held the positions of manager and chief director. His interests are focused in the fields of Distributed Systems, Big Data, Simulation, and Natural Language Processing.

An Entrepreneurial Well-being Model based on GEM Data for Spain

José Manuel Saiz-Alvarez[1], Alicia Coduras Martínez[1,2], and Carlos Cuervo-Arango Martínez[1]

[1]*Nebrija University*
[2]*GEM Global*

***Abstract* — The Economics of Happiness is one of the research areas of greatest growth in recent years. Throughout this work, a venture based model in which satisfaction of Spanish entrepreneurs with their professional life is performed. We analyze the responses of 9,989 entrepreneurs using data from the Global Entrepreneurship Monitor (GEM), and six hypothesis are discussed. The results show that, for the Spanish case, there is a strong consistency in the results the opportunity entrepreneurs present greater satisfaction than necessity entrepreneurs.**

***Keywords* — Well-being, education, entrepreneurship, happiness.**

I. Introduction

THE economics of happiness is the theoretical and quantitative study of the effects on well-being, or other subjective measures of life satisfaction of different circumstances or events which have a determined impact on the life´s conditions of individuals. It relates subjective measures of satisfaction with objectives "states of the world"; the purpose being to extract conclusions about the influence that the later could have on the perceived level of satisfaction of those affected.

As shown by Ferrer-i-Carbonell [11] happiness or relative subjectivity can be constituted as a proxy measure of utility to better understand the preferences of individuals in relation to issues as diverse as revenue received in connection with the reference group, working conditions, unemployment, health and socioeconomic inequality. Thus, for instance, with regard to subjective well-being, women are happier than men and their happiness varies with age according to an "U" shape: between 25 and 40 years happiness decreases until reaching a point between 40 and 50 years when happiness starts growing, as stated by Guardiola [13]. This female behavior is generally linked to strong family ties. As the *World Values Survey* and the *Gallup Survey* show, appropriate wage remuneration for work done and good relationships with family and friends are the most influential variables to increase the relative well-being or happiness.

On a different track, after having tested more than 20,000 participants, MacKerron & Mourato [16] show that people is significantly and substantially happier outdoors in all green or natural habitat types than if they live in urban environments, same conclusion reached by Deschacht [9]. This fact allows Cuñado & Pérez de Gracia [7] to compute the monetary value of air quality and climate, deriving the average marginal rate of substitution between income, air quality and climate for the Spanish regions. In fact, Ferrer-i-Carbonell and Gowdy [12] associate unhappiness with environmental degradation, which encouraged the government's implementation of policies focused on improving the environment.

Finally, happiness is also affected by experiential and material purchases. While experiential purchases serve the purpose of acquiring a life experience, material purchases are focused on acquiring an object. Thomas & Millar [21] show that experiential purchases are associated with more happiness than material purchases.

However, the consideration given to entrepreneurship as a determinant of well-being or happiness, has received much less attention in the literature, the fact notwithstanding that there are various arguments to speak for the self-stem, appreciation and sense of fulfillment that those involved obtain from the very act of developing a new project or creating a new firm from scratch.

There is thus, a basis to consider entrepreneurship as a key factor to increase social wealth, by way of increasing the happiness or well-being of the people concerned. The goal of this paper is to study the relationship among entrepreneurship and well-being using GEM data for Spain. To cope with this objective, we shall begin characterizing the relation between entrepreneurship and happiness to lay the foundations of a model relating both variables.

II. Entrepreneurship and Happiness

In a very early study, Clark, Colombier & Nasclet [4] analyze whether self-employed entrepreneurs are more satisfied than salaried workers, and using French and English data find that first-generation self-employed (those whose parents were not self-employed) are more satisfied overall than are the second-generation self-employed.

The act of entrepreneurship typically confers a trade-off between distinctiveness and the risk of diminishing psychological well-being, as in Shepherd & Haynie [18]. To be different from other professionals, entrepreneurs take commercial and financial risks. Those risks have to be

.

DOI: 10.9781/ijimai.2014.255

measured carefully making the necessary provisions for a proper functioning of the company. Failure to do so will augment the risk of failure, leading to dissatisfaction and a sense of personal failure reinforced if other entrepreneurs achieve success in their business.

After having analyzed data from 156 entrepreneurs, Uy, Foo & Song [22] indicate that successfully implementing new ideas is linked with more happiness when starting-up a company. This feeling is more intense in opportunity rather than necessity entrepreneurs. As a result, when new ideas are successfully being put into action, more new ideas are being generated and a virtuous circle results.

When crisis is deep, the apparent financial irrationality of entrepreneurship is typically explained in terms of non-pecuniary compensation factors, such as autonomy and satisfaction, as shown by Carter [3]. In fact, a large percentage of entrepreneurs create businesses after years of having been employed. As a result, they have the training, experience and desire to be entrepreneurs, so they have the strength to overcome any obstacles they may face in their path to success. The greater the number of obstacles to overcome on the road to success, the greater the feeling of happiness once entrepreneurial success is achieved.

Binder & Coad [2] find that individuals who move from regular employment into self-employment experience an increase in life satisfaction up to two years later, while individuals moving from unemployment to self-employment are not more satisfied than their counterparts moving from unemployment to regular employment. Unemployment is an undesirable situation by reducing the individual's self-esteem, well-being and happiness. Therefore, both the probability of entering unemployment in Origoa & Paganib [17], and temporary contracts, as shown by Kaiser [14] are negatively correlated with job satisfaction. As a result, entrepreneurial rewards are not only determined by business experience, formation and rationality, but are influenced by changing needs over time in Carter [3] which, when satisfied by the entrepreneur, increases his/her self-esteem and sense of accomplishment.

In a very different latitude, Cortés, García & Moro-Egido [6] study the relationship between labor status and individual satisfaction in Latin America. To clarify the effect of self-employment on satisfaction, they use the *Latinobarómetro Survey 2007* for eighteen Latin American and Caribbean countries, considering the category self-employment as a heterogeneous category. Contrary to existing evidence, they find that not all self-employed individuals are more satisfied than generally employed people. However, controlling for the distinction between necessity and opportunity self-employment, they obtain positive associations between self-employed entrepreneurs and their subjective well-being, although the relation does not extend across all categories of self-employed workers.

The international scientific community had achieved a certain consensus about the limited role of economic growth on the generation of happiness in the long-term as stated by Easterlin [10], and on the acceptance of the so-called Easterlin Paradox (Substantial increments in income are not accompanied by increases in the levels of reported happiness). However, Stevenson & Wolfers [20] reopened the debate to find a small positive relationship between the two variables in the United States and in some European countries. This relationship between the desire to undertake and the feeling of happiness must be particularly intense in entrepreneurs, which makes them into internal drivers for change in organizations.

Lofstrom [15] finds that, although the returns of low-skilled self-employment among men is higher than among women, wage/salary employment is a more financially rewarding option for most low-skilled workers. Despite this fact, low-skilled workers do not fare well in today's skill intensive economy and their working opportunities continue to diminish.

The entrepreneur judged their quality of life in relative terms to compare their achievements with a group of people nearby, whether family, friends or neighbors. Therefore, as shown by Ferrer-i-Carbonell [11] economic growth distributed equally has little impact on the reported happiness. Therefore, according to Van Praag [23], to measure the degree of happiness the existence of a reference group (set-point theory) must be taken into account, so it can serve as a basis for measuring subjective feelings about perceived inequality by individuals, being a temporary feeling with continuous fluctuations in happiness (hedonic treadmill). In extreme comparing situations, individual suicide risk rises with others' income, as demonstrated by Daly, Wilson & Johnson [8]. But in these cases, psychological and psychiatric factors, frequently too complex, affect more than purely economic ones.

Entrepreneurs can be classified by a number of psychological characteristics defined by what we call the entrepreneurial spirit (see Figure 1). Ambition leads to outstanding performance, and the latter derives largely from participation in prolonged, intense, and highly-focused efforts to improve current performance, defined by deliberate practice, as shown in Baron & Henry [1].

III. Model and Research Hypotheses

Given the above mentioned premises, we ask the following general research question: Does entrepreneurship make a difference in the satisfaction with their professional life of adult Spaniards?

We start by questioning the ability of different people to acquire their own skills by themselves, thus valuing them as achievements that make them capable to reach levels of achievement that they could not attain otherwise. Even though this level of effort could mean an increased level of personal tension and even stress, it also allows higher levels of income. These, by themselves, increase the perceived level of satisfaction of the entrepreneurs involved and translate also into a general perception of well-being that extends beyond the pure salary conditions to a feeling of overall satisfaction.

Within this framework, we tested the following hypothesis:
H1: The segment of the population involved in entrepreneurial activities is more capable of developing their own skills than the segment of the population not involved.
H2: The segment of the population involved in entrepreneurial activities is more prone to believe the work they do is meaningful than the segment of the population not involved.
H3: The segment of the population involved in entrepreneurial activities is more exposed to excessive stress than the segment of the population not involved.

Considering the results obtained by Clark, Colombier & Nasclet [4], we tested the previous hypotheses for different segmentations of the population based in the length of their experience in the entrepreneurial environment:
H4: The segment of the population involved in entrepreneurial activities is overall more professionally satisfied than the segment of the population not involved.
H5: The segment of the population involved in entrepreneurial activities is overall more satisfied with their current work income than the segment of the population not involved.

Finally, we tested the following complementary hypothesis to see if the results defended by Uy, Foo & Song [22] about the influence of necessity and opportunity entrepreneurship could be reproduced for Spain or, on the contrary, the situation is closer to the results obtained by Cortés, García & Moro-Egido [6]:
H6: Early stage entrepreneurs by necessity are less happy or satisfied with their professional life than opportunity early stage entrepreneurs

IV. Methodology and Empirical Validation

As explained in the introduction, this research is focused on Spain. The data to answer the general research question and test the associated hypotheses were provided by the GEM adult population survey data base for Spain 2012. This data base includes observations of 9,989 individuals who acted as respondents of key items on entrepreneurial attitudes, aspirations and activity along with some items on professional well-being. The dependent variables included in the Spanish GEM data base were those showed in Table 1.

Table 1
Key items on well-being included in the GEM Spanish survey for the year 2012

1	I can decide on my own how I go about doing my work
2	The work I do is meaningful to me
3	At my work, I am not exposed to excessive stress
Scale: Likert, 5 points from 1 = strongly disagree to 5 = strongly agree	
4	Overall, how satisfied are you with your current work?
5	Overall, how satisfied are you with your current work income?
Scale: Likert, 5 points from 1 = very dissatisfied to 5 = very satisfied	

In accordance with our model, we selected as dependent variables various indicators of the "well-being/satisfaction with professional life" and as independent variable "the entrepreneurial status of an individual". In other words, we made the general hypothesis that the well-being/satisfaction with professional life varies depending on some indicators of the degree of advance in the way of an entrepreneurial status.

All these variables were qualitative and ordinal, condition that determined the selection of a proper analysis method. In this case, a nonparametric test for multiple independent samples was used to test whether or not the values of a particular variable differed between two or more groups of individuals when the assumptions of ANOVA are not met, which is the current situation.

As is well known, although one-way analysis of variance (ANOVA) is the method of choice when testing for differences between multiple groups, it assumes that the mean is a valid estimate of center and that the distribution of the test variable is reasonably Normal and similar in all groups. However, in our case the behavior of the dependent variables ratifies that the ANOVA conditions are not met (see results in Table 2), since none of them is normally distributed.

Table 2
Normality tests on key items on well-being included in the GEM Spanish survey for the year 2012

Criteria: the variable is Normal when Significance is greater than 0.025	Kolmogorov-Smirnov tests:			
Items:	Mean	Std. Dv.	Z K-S	Sig. (Bilateral)
I can decide on my own how I go about doing my work	3.7457	1.287	29.088	0.000
The work I do is meaningful to me	4.4786	0.857	35.582	0.000
At my work, I am not exposed to excessive stress	2.7336	1.452	22.520	0.000
Overall, how satisfied are you with your current work?	4.0120	0.891	33.120	0.000
Overall, how satisfied are you with your current work income?	3.1469	1.144	27.772	0.000

When the assumptions behind the standard ANOVA are invalid, one should consider using the nonparametric procedures designed to test for the significance of the difference between multiple groups. They are called nonparametric because they make no assumptions about the parameters of the distribution, nor do they assume that any particular distribution is being used, as shows Conover [5]. For our analysis we selected the Median Test.

The median method tests the null hypothesis that two or more independent samples have the same median, as in Siegel & Castellan [19]. In our case we tested that people involved and not involved in entrepreneurial activities have the same

median value for the five key items measuring professional satisfaction/well-being. This test assumes nothing about the distribution of the test variables, making it a good choice when one suspects that the distribution varies by group.

To apply these tests, our first need was to establish the entrepreneurial status of the individuals, that is, the independent variable. For this purpose, we classified the respondents of the GEM survey in two groups as follows: [1] Persons involved in entrepreneurial activities, and [2] Persons not involved in entrepreneurial activities.

However, to define the entrepreneurial status of an individual is not as easy as it seems because it is possible to consider different ways of grouping. Thus, within the GEM Project, a person can be: [1] A potential entrepreneur: person that indicates entrepreneurial intention; [2] A nascent entrepreneur: person involved in a startup of no more than 3 months in the market; [3] A new entrepreneur: person involved in consolidating a new business with no more than 3.5 years in the market; [4] An established entrepreneur: owner-manager of a consolidated business operating more than 3.5 years in the market, and [5] An exited entrepreneur: person that abandoned-exited an activity within the 12 months period previous to the survey.

Thus, one possibility was to create a dichotomous variable including all these cases in the group of involved people and the rest in the category of not involved. Another was to discard the exited entrepreneurs because one can suspect that their sense of happiness could be different from those that are active entrepreneurs. Another was to discard the exited entrepreneurs and also the potential entrepreneurs, as it can be expected that they have not yet experienced the same feelings as active entrepreneurs. And, finally, another possibility was to include only early-stage entrepreneurs in the category of involved, that is, nascent and new entrepreneurs, and put the rest on the category of not involved.

It was tempting to explore what could be the results for each grouping and we yielded building four models of dependent variable to test all possibilities. The wide sample provided by the data base gave support to develop this exercise as there were enough cases (more than 400 per subsample) for the four grouping models. The decision taken resulted in building the set of dichotomous, independent variables described in Table 3.

Summarizing, we applied Median Tests to draw conclusions on the following hypotheses, now expressed in statistical language:

H1: median agreement on the statement "I can decide on my own how I go about doing my work" is equal between population involved and not involved in entrepreneurial activities.

H2: median agreement on the statement "the work I do is meaningful to me" is equal between population involved and not involved in entrepreneurial activities.

H3: median agreement on the statement "at my work, I am not exposed to excessive stress" is equal between population involved and not involved in entrepreneurial activities.

H4: median satisfaction with current work is equal between population involved and not involved in entrepreneurial activities.

H5: median satisfaction with current work income is equal between population involved and not involved in entrepreneurial activities.

H6: the median scores of all the professional well-being items are significantly higher for opportunity early stage entrepreneurs compared with the scores of necessity early stage entrepreneurs.

TABLE 3
VARIABLES USED TO ESTABLISH THE CONTRAST GROUPS

1. Entrepreneurial status of individuals in the sample (Wide group):		Values
Involved	Potential, nascent, new, established and exited entrepreneurs	1
Not involved	Employees part or full time, students, home makers, unemployed, retired	0
2. Entrepreneurial status of individuals in the sample (Less wide):		Values
Involved	Potential, nascent, new, and established entrepreneurs	1
Not involved	Employees part or full time, students, home makers, unemployed, exited entrepreneurs, retired	0
3. Entrepreneurial status of individuals in the sample (Only active):		Values
Involved	Nascent, new, and established entrepreneurs	1
Not involved	Potential entrepreneurs, employees part or full time, students, home makers, unemployed, exited entrepreneurs, retired	0
4. Entrepreneurial status of individuals in the sample (Only early stage)		Values
Involved	Nascent and new entrepreneurs	1
Not involved	Potential entrepreneurs, established businesses owner-managers, employees part or full time, students, home makers, unemployed, exited entrepreneurs, retired	0
5. Main motive for starting up a business (Only early stage)		Values
Opportunity	Early stage entrepreneurs that started up pursuing an opportunity	1
Necessity	Early stage entrepreneurs that started up under the lack of better work options	0

The results for these tests are presented in the next section.

V. RESULTS

Entrepreneurial status including potential, nascent, new, established and exited entrepreneurs

In this case, we considered that people involved in entrepreneurial activities were either, potential, nascent, new, established or exited entrepreneurs and people not involved the rest of the labor categories. The results obtained for the median test are summarized and showed in the following tables.

Across all subjects, the medians are scores of 4, 5, 2, 4 and 4 points. The null hypothesis for the median test is that these particular values are good approximations of center for each of the two groups for each independent variable. The table 6 shows the results of these tests for each independent variable.

TABLE 4
MEDIAN TEST, DESCRIPTIVE RESULTS

Dependent variables	n	Quartiles		
		Q1 (25%)	Q2 (50%)	Q3 (75%)
I can decide on my own how I go about doing my work	9712	3(Neither agree nor disagree)	4(somewhat agree)	5 (strongly agree)
The work I do is meaningful to me	9718	4(somewhat agree)	5(strongly agree)	5 (strongly agree)
At my work, I am not exposed to excessive stress	9712	1(strongly disagree)	2(somewhat disagree)	4(somewhat agree)
Overall, how satisfied are you with your current work?	9731	4(satisfied)	4(satisfied)	5(very satisfied)
Overall, how satisfied are you with your current work income?	9698	2(unsatisfied)	4(satisfied)	4(satisfied)
Independent variable:	Involved or not in entrepreneurial activities (1 = potential, nascent, new, established, exited; 0 = rest)			

TABLE 5
MEDIAN TEST, DISTRIBUTIONS
(PERCENTAGES OF CASES BELOW/EQUAL THE MEDIAN OR OVER THE MEDIAN)

Independent variable: involved or not in entrepreneurship		NO	YES
I can decide on my own how I go about doing my work (the median was 4 points: somewhat agree)	>Median ≤Median	28.9% **71.1%**	**50.4%** 49.6%
The work I do is meaningful to me (the median was 5 points: strongly agree)	>Median ≤Median	0.0% 100.0%	0.0% 100.0%
At my work, I am not exposed to excessive stress (the median was 2: somewhat disagree)	>Median ≤Median	46.5% **53.5%**	46.4% **53.6%**
Overall, how satisfied are you with your current work? (the median was 4: satisfied)	>Median ≤Median	27.0% **73.0%**	30.6% **69.4%**
Overall, how satisfied are you with your current work income? (the median was 4: satisfied)	>Median ≤Median	7.2% **92.8%**	7.9% **92.1%**

TABLE 6
CONTRAST STATISTICS

Dependent variables:	Chi Sq.	As. Sig.	Hypothesis
I can decide on my own how I go about doing my work (The median was 4 points: somewhat agree)	342.3	**0.000**	**Rejected**
The work I do is meaningful to me (The median was 5 points: strongly agree and all values were under or equal the median)	Not calculable	Not calculable	Accepted
At my work, I am not exposed to excessive stress (The median was 2: somewhat disagree)	0.000	0.985	Accepted
Overall, how satisfied are you with your current work? (The median was 4: satisfied)	10.4	**0.001**	**Rejected**
Overall, how satisfied are you with your current work income? (The median was 4: satisfied)	1.39	0.238	Accepted

Entrepreneurial status including potential, nascent, new and established entrepreneurs

In this case, we considered that people involved in entrepreneurial activities were either, potential, nascent, new and established entrepreneurs and people not involved the rest of labor categories including now exited entrepreneurs.

The results obtained for the median test are summarized and showed in the following tables.

Across all subjects, the medians are scores of 4, 5, 2, 4 and 4 points. The null hypothesis for the median test is that these particular values are good approximations of center for each of the two groups for each independent variable. The table 7 shows the results of these tests for each independent variable.

TABLE 7
MEDIAN TEST, DESCRIPTIVE RESULTS

Dependent variables	n	Quartiles		
		Q1 (25%)	Q2 (50%)	Q3 (75%)
I can decide on my own how I go about doing my work	9712	3(Neither agree nor disagree)	4(somewhat agree)	5(strongly agree)
The work I do is meaningful to me	9718	4(somewhat agree)	5(strongly agree)	5(strongly agree)
At my work, I am not exposed to excessive stress	9712	1(strongly disagree)	2(somewhat disagree)	4(somewhat agree)
Overall, how satisfied are you with your current work?	9731	4(satisfied)	4(satisfied)	5(very satisfied)
Overall, how satisfied are you with your current work income?	9698	2(unsatisfied)	4(satisfied)	4(satisfied)
Independent variable:	Involved or not in entrepreneurial activities (1 = potential, nascent, new, established; 0 = rest)			

TABLE 8
MEDIAN TEST, DISTRIBUTIONS
(PERCENTAGES OF CASES BELOW/EQUAL THE MEDIAN OR OVER THE MEDIAN)

Independent variable: involved or not in e-ship--→		NO	YES
I can decide on my own how I go about doing my work (the median was 4 points: somewhat agree)	>Median ≤Median	29.0% **71.0%**	**51.2%** 48.8%
The work I do is meaningful to me (the median was 5 points: strongly agree)	>Median ≤Median	0.0% 100.0%	0.0% 100.0%

At my work, I am not exposed to excessive stress (the median was 2: somewhat disagree)	>Median ≤Median	46.6% **53.4%**	46.0% **54.0%**
Overall, how satisfied are you with your current work? (the median was 4: satisfied)	>Median ≤Median	36.9% **73.1%**	31.0% **69.0%**
Overall, how satisfied are you with your current work income? (the median was 4: satisfied)	>Median ≤Median	7.2% **92.8%**	7.8% **92.2%**

TABLE 9
CONTRAST STATISTICS

Dependent variables:	Chi Sq.	As. Sig.	Hypothesis
I can decide on my own how I go about doing my work (the median was 4 points: somewhat agree)	351.5	**0.000**	**Rejected**
The work I do is meaningful to me (the median was 5 points: strongly agree and all values were under or equal the median)	Not calculable	Not calculable	Accepted
At my work, I am not exposed to excessive stress (the median was 2: somewhat disagree)	0.148	0.700	Accepted
Overall, how satisfied are you with your current work? (the median was 4: satisfied)	12.9	**0.000**	**Rejected**
Overall, how satisfied are you with your current work income? (the median was 4: satisfied)	1.39	0.804	Accepted

Entrepreneurial status including nascent, new and established entrepreneurs

In this case, we considered that people involved in entrepreneurial activities were either, nascent, new and established entrepreneurs and people not involved the rest of labor categories including now exited entrepreneurs and potential entrepreneurs.

The results obtained for the median test are summarized and showed in the following tables.

Across all subjects, the medians are scores of 4, 5, 2, 4 and 4 points. The null hypothesis for the median test is that these particular values are good approximations of center for each of the two groups for each independent variable. The table 10 shows the results of these tests for each independent variable.

TABLE 10
MEDIAN TEST, DESCRIPTIVE RESULTS

Dependent variables	**n**	**Quartiles**		
		Q1 (25%)	**Q2 (50%)**	**Q3 (75%)**
I can decide on my own how I go about doing my work	9712	3(Neither agree nor disagree)	4(somewhat agree)	5 (strongly agree)
The work I do is meaningful to me	9718	4(somewhat agree)	5(strongly agree)	5 (strongly agree)
At my work, I am not exposed to excessive stress	9712	1(strongly disagree)	2(somewhat disagree)	4(somewhat agree)
Overall, how satisfied are you with your current work?	9731	4(satisfied)	4(satisfied)	5(very satisfied)
Overall, how satisfied are you with your current work income?	9698	2(unsatisfied)	4(satisfied)	4(satisfied)
Independent variable:	Involved or not in entrepreneurial activities (1 = nascent, new, established; 0 = rest)			

TABLE 11
MEDIAN TEST, DISTRIBUTIONS
(PERCENTAGES OF CASES BELOW/EQUAL THE MEDIAN OR OVER THE MEDIAN)

Independent variable: involved or not in e-ship--→		NO	YES
I can decide on my own how I go about doing my work (the median was 4 points: somewhat agree)	>Median ≤Median	32.0% **68.0%**	**43.2%** 56.8%
The work I do is meaningful to me (the median was 5 points: strongly agree)	>Median ≤Median	0.0% 100.0%	0.0% 100.0%
At my work, I am not exposed to excessive stress (the median was 2: somewhat disagree)	>Median ≤Median	46.5% **53.5%**	46.0% **54.0%**
Overall, how satisfied are you with your current work? (the median was 4: satisfied)	>Median ≤Median	27.5% **72.5%**	29.5% **70.5%**
Overall, how satisfied are you with your current work income? (the median was 4: satisfied)	>Median ≤Median	7.2% **92.8%**	8.6% **91.4%**

TABLE 12
CONTRAST STATISTICS

Dependent variables:	Chi Sq.	As. Sig.	Hypothesis
I can decide on my own how I go about doing my work (the median was 4 points: somewhat agree)	64.6	**0.000**	**Rejected**
The work I do is meaningful to me (the median was 5 points: strongly agree and all values were under or equal the median)	Not calculable	Not calculable	Accepted
At my work, I am not exposed to excessive stress (the median was 2: somewhat disagree)	0.097	0.756	Accepted
Overall, how satisfied are you with your current work? (the median was 4: satisfied)	2.4	**0.121**	**Accepted**
Overall, how satisfied are you with your current work income? (the median was 4: satisfied)	3.5	**0.061**	**Accepted at 95% Rejected at 90%**

Entrepreneurial status including nascent and new entrepreneurs (people involved in early stage entrepreneurial activities)

In this case, we considered that people involved in entrepreneurial activities were only nascent and new entrepreneurs and people not involved the rest of labor categories including now exited entrepreneurs, potential and established entrepreneurs.

The results obtained for the median test are summarized and showed in the following tables.

Across all subjects, the medians are scores of 4, 5, 2, 4 and 4 points. The null hypothesis for the median test is that these particular values are good approximations of center for each of the two groups for each independent variable. The table 13 shows the results of these tests for each independent variable.

TABLE 13
MEDIAN TEST, DESCRIPTIVE RESULTS

Dependent variables	**n**	**Quartiles**		
		Q1 (25%)	**Q2 (50%)**	**Q3 (75%)**
I can decide on my own how I go about doing my work	9712	3(neither agree nor disagree)	4(somewhat agree)	5 (strongly agree)
The work I do is meaningful to me	9718	4(somewhat agree)	5(strongly agree)	5 (strongly agree)
At my work, I am not exposed to excessive stress	9712	1(strongly disagree)	2(somewhat disagree)	4(somewhat agree)
Overall, how satisfied are you with your current work?	9731	4(satisfied)	4(satisfied)	5(very satisfied)
Overall, how satisfied are you with your current work income?	9698	2(unsatisfied)	4(satisfied)	4(satisfied)
Independent variable:	Involved or not in entrepreneurial activities (1 = nascent, new; 0 = rest)			

TABLE 14
MEDIAN TEST, DISTRIBUTIONS
(PERCENTAGES OF CASES BELOW/EQUAL THE MEDIAN OR OVER THE MEDIAN)

Independent variable: involved or not in entrepreneurship		NO	YES
I can decide on my own how I go about doing my work (the median was 4 points: somewhat agree)	>Median ≤Median	33.2% **66.8%**	**37.2%** 62.8%
The work I do is meaningful to me (the median was 5 points: strongly agree)	>Median ≤Median	0.0% 100.0%	0.0% 100.0%
At my work, I am not exposed to excessive stress (the median was 2: somewhat disagree)	>Median ≤Median	46.5% **53.5%**	46.0% **54.0%**
Overall, how satisfied are you with your current work? (the median was 4: satisfied)	>Median ≤Median	27.9% **72.1%**	26.3% **73.7%**
Overall, how satisfied are you with your current work income? (the median was 4: satisfied)	>Median ≤Median	7.1% **92.9%**	9.2% **90.8%**

TABLE 15
CONTRAST STATISTICS

Dependent variables:	Chi Sq.	As. Sig.	Hypothesis
I can decide on my own how I go about doing my work (the median was 4 points: somewhat agree)	6.3	**0.012**	**Rejected**
The work I do is meaningful to me (the median was 5 points: strongly agree and all values were under or equal the median)	Not calculable	Not calculable	Accepted
At my work, I am not exposed to excessive stress (the median was 2: somewhat disagree)	0.094	0.759	Accepted
Overall, how satisfied are you with your current work? (the median was 4: satisfied)	1.143	0.285	Accepted
Overall, how satisfied are you with your current work income? (the median was 4: satisfied)	4.9	**0.026**	**Rejected**

Opportunity versus necessity early stage entrepreneurs

TABLE 16
MEDIAN TEST, DESCRIPTIVE RESULTS

Dependent variables	**n**	**Quartiles**		
		Q1 (25%)	**Q2 (50%)**	**Q3 (75%)**
I can decide on my own how I go about doing my work	712	4(somewhat agree)	5(strongly agree)	5 (strongly agree)
The work I do is meaningful to me	710	5(strongly agree)	5(strongly agree)	5 (strongly agree)
At my work, I am not exposed to excessive stress	710	1(strongly disagree)	2(somewhat disagree)	4(somewhat agree)
Overall, how satisfied are you with your current work?	712	4(satisfied)	4(satisfied)	5(very satisfied)
Overall, how satisfied are you with your current work income?	702	2(unsatisfied)	3(neither agree nor disagree)	4(satisfied)
Independent variable:	Motive (1 = opportunity; 0 = necessity)			

TABLE 17
CONTRAST STATISTICS

Dependent variables:	Chi Sq.	As. Sig.	Hypothesis
I can decide on my own how I go about doing my work (the median was 4 points: somewhat agree)	Not calculable	Not calculable	Accepted
The work I do is meaningful to me (the median was 5 points: strongly agree and all	Not calculable	Not calculable	Accepted

values were under or equal the median)			
At my work, I am not exposed to excessive stress (the median was 2: somewhat disagree)	0.898	0.343	Accepted
Overall, how satisfied are you with your current work? (the median was 4: satisfied)	1.396	0.237	Accepted
Overall, how satisfied are you with your current work income? (the median was 4: satisfied)	16.4	**0.000**	**Rejected**

VI. Conclusions

Analysis considering entrepreneurial status including potential, nascent, new, established and exited entrepreneurs

As result of the first analysis two hypotheses were rejected and three accepted. Thus, when the entrepreneurial condition includes potential, nascent, new, established and exited entrepreneurs: [1] the median agreement on the statement "I can decide on my own how I go about doing my work" is significantly different between population involved and not involved in entrepreneurial activities. In this case, the mean is greater for the involved population; [2] the median agreement on the statement "the work I do is meaningful to me" is equal between population involved and not involved in entrepreneurial activities. In this case, the entrepreneurial status makes no difference, at least in Spain; [3] the median agreement on the statement "at my work, I am not exposed to excessive stress" is equal between population involved and not involved in entrepreneurial activities. So, in this case the entrepreneurial status makes no difference in Spain; [4] the median satisfaction with current work is different between population involved and not involved in entrepreneurial activities. In this case people involved in entrepreneurial activities are more satisfied than people not involved, and [5] the median satisfaction with current work income is equal between population involved and not involved in entrepreneurial activities. In Spain, despite the crisis, entrepreneurs are more prone to decide on their own how they go about doing their work and derive from it a higher degree of satisfaction with their work. Accordingly to the literature, this independent behavior could result in a greater feeling of professional happiness. Thus, entrepreneurship makes some difference at least in these aspects. The measurement was done just in the middle of the crisis so all the population seems to be affected by a profound stress that probably overcomes any possible effect due to the different nature of the occupation of the two population groups. The same conclusion could be drawn regarding the satisfaction with work income. Finally, Spaniards point out that they are generally happy with their jobs and they are very meaningful for more than one half of the population, no matter whether they are entrepreneurs or not.

Analysis considering entrepreneurial status formed by potential, nascent, new, and established entrepreneurs

The results are very similar to those obtained in the previous analysis. There is a difference in that the strength of the diagnoses of acceptance/ rejection of hypothesis in this analysis is more extreme. The conclusions drawn before remain the same and removing exited entrepreneurs from the group of those involved in entrepreneurial activities makes no relevant difference.

Analysis considering entrepreneurial status formed by nascent, new, and established entrepreneurs

The results are similar to those obtained in the previous analyses for three dependent variables and different for two of them. Thus, the hypothesis about the item "Overall, how satisfied are you with your current work?" is now accepted while it was rejected before. This means that entrepreneurship is understood as a status including nascent, new and established entrepreneurs makes no difference in Spain (in the middle of a crude economic crisis) in relation from the satisfaction derived from the current job. On the contrary, now the hypothesis on the item "Overall, how satisfied are you with your current work income?" tends to be rejected while it was accepted in the previous considerations on the entrepreneurial status. This means that nascent, new and established entrepreneurs as a group tend to be slightly more satisfied than the rest of the population about their current work income (see Table 11).

Finally it can be also noticed a difference in the strength of the diagnoses of acceptance and rejection of hypotheses, which is more extreme.

Analysis considering entrepreneurial status formed by nascent and new entrepreneurs

The results are similar to the previous ones except by two facts. One is that the hypothesis about the item "Overall, how satisfied are you with your current work income?" is now more clearly rejected than before, and second, that the hypothesis on the item "I can decide on my own how I go about doing my work" is now less clearly rejected. This provides more strong evidence supporting that early stage entrepreneurs are more satisfied than the rest of the categories with their work income, but less strong evidence supporting that this group is more prone to consider that they can decide on their own how they go about doing their work.

The obtained results give some support to the idea that entrepreneurs tend to be more satisfied or happy with their professional lives and tend to ratify Carter's [3] theories. However, it is apparent that the differences between the groups involved and not involved in entrepreneurial activities are far from being extreme in Spain. The aspects which appear as differential are the capability of controlling how people goes about doing their work, the satisfaction with it (depending on

the categories included in the entrepreneurial status), and the satisfaction with the income derived depending on the categories included in the entrepreneurial status.

Analysis for opportunity and necessity early stage entrepreneurs

The previous results induce the consideration of other variables involved in the determination of the degree of happiness or satisfaction derived by individuals from their professional life. Among them, the distinction between opportunity and necessity entrepreneurship takes pride of place. Nevertheless, the results for Spanish early stage entrepreneurs only give support to the idea that satisfaction with current work income is higher among opportunity entrepreneurs. The rest of items reach similar median scores, and there are no significant differences between opportunity and necessity early stage entrepreneurs.

This points out that in these regards, the entrepreneurial collective in Spain is more similar to Latin-America, as described by Cortés, García & Moro-Egido [6], and it does not show the characteristics defended by Uy, Foo & Song [22].

Policy implications, final remarks and next research lines

Other tests made over the GEM Spanish sample have given no significant results regarding gender differences, fear of failure or having or not completed specific training to startup businesses and other relevant variables.

This overall result can be interpreted as some exceptional because it is not aligned with the main conclusions derived from specialized literature. The economic crisis can be an explanatory factor of the major homogeneity of adult Spaniards with respect to professional well-being but, at the same time, it is important to point out that we, as researchers, are confronting a structural change in developed societies.

If the impulse of micro and SMEs progressively substitutes the economy based in big companies, the social panorama will change dramatically and the professional well-being parameters can also change substantially. As some authors pointed out, the measurement of well-being depends on a good determination of a reference point, and it seems that it is in the middle of significant structural changes.

In Spain and in other countries, the policy makers must take all of this into consideration as the policies' design will confront (is confronting now) important challenges to overcome unemployment rates, youth unemployment, and the dissatisfaction of a big part of the population that considers is quickly losing their quality of life. In such scenario, entrepreneurship can become more a professional than a vocational option and, due to this, it is possible that it is losing some of its associated "fresh" characteristics.

Entrepreneurial education is progressing and it is time to reflect on what kind of model the university must convey to students to keep the most important values of entrepreneurship. Future lines of research should focus on comparing the Spanish well-being indicators in the entrepreneurial and non-entrepreneurial population of the next years with the current results, in order to determine if the background is changing and in what degree.

REFERENCES

[1] Baron, R.A. & Henry. R.A., "How entrepreneurs acquire the capacity to excel: Insights from research on expert performance". *Strategic Entrepreneurship Journal*. vol. 4, n. 1, 2010, pp. 49-65.

[2] Binder, M. & Coad, A., "Life Satisfaction and Self-employment: A Matching Approach", *Small Business Economics*, vol. 40, n. 4, 2013, pp. 1009-1033.

[3] Carter, S., "The Rewards of Entrepreneurship: Exploring the Incomes, Wealth, and Economic Well-Being of Entrepreneurial Households", *Entrepreneurship. Theory and Practice*, vol. 35, n. 1, 2011, pp. 39-55.

[4] Clark, A.E.; Colombier, N. & Masclet, D., "Never the same after the first time: the satisfaction of the second-generation self-employed", IZA Discussion Papers, n. 3476, Bonn (Germany): Institute for the Study of Labor (IZA), 2008.

[5] Conover, W.J., *Practical Nonparametric Statistics*, 2nd ed. New York: John Wiley and Sons, 1980.

[6] Cortés, A.; García, T. & Moro-Egido, A.I., "Heterogeneous Self-employment and Satisfaction in Latin America", *Journal of Economic Psychology*, vol. 39, 2013, p. 44-61.

[7] Cuñado, J. & Pérez de Gracia, F., "Environment and Happiness: New Evidence for Spain", *Social Indicators Research*, vol. 112, n. 3, 2013, pp. 549-567.

[8] Daly, M.C.; Wilson, D.J & Johnson, N.J., "Relative Status and Well-being: Evidence from U.S. Suicide Deaths", *Review of Economics and Statistics*, vol. 95, n. 5, 2013, pp. 1480-1500.

[9] Deschacht, N., "The Economics of Happiness: The Conservative Current in Green Economic Thought", *Science & Society*, vol. 77, n. 4, 2013, pp. 569-575.

[10] Easterlin, R.A., "Does economic growth improve the human lot? Some empirical evidence", in David, P.A. & Reder, M.W. (Eds.): *Nations and Households in Economic Growth. Essays in Honor of Moses Abramowitz*, New York: Academic Press, 1974, pp. 89-125.

[11] Ferrer-i-Carbonell, A., "Economía de la Felicidad", *Els Opuscles del CREI*, n. 28, may, Barcelona (Spain): Universitat Pompeu Fabra, 2011.

[12] —. & Gowdy, J.M., "Environmental degradation and happiness", *Ecological Economics*, Elsevier, vol. 60(3), 2007, pp. 509-516.

[13] Guardiola, J., "¿Qué aportan los estudios de felicidad al buen vivir y viceversa?", *Obets. Revista de Ciencias Sociales*, vol. 6, n. 1, 2011, pp. 97-109.

[14] Kaiser, L.C., "Job satisfaction: A comparison of standard, non-standard, and self-employment patterns across Europe with a special note to the gender/job satisfaction paradox", EPAG Working Paper No. 27, 2002.

[15] Lofstrom, M., "Does self-employment increase the economic well-being of low-skilled workers?", *Small Business Economics*, vol. 40, n. 4, 2013, pp. 933-952.

[16] MacKerron, G. & Mourato, S., "Happiness is greater in natural environments", *Global Environmental Change-Human and Policy Dimensions*, vol. 23, n. 5, 2013, pp. 992-1000.

[17] Origoa, F. & Paganib, L., "Flexicurity and job satisfaction in Europe: The importance of perceived and actual job stability for well-being at work", *Labour Economics*, n. 16, 2009, pp. 547-555.

[18] Shepherd, D. & Haynie, J.M., "Birds of a feather don't always flock together: Identity Management in Entrepreneurship", *Journal of Business Venturing*, vol. 24, n. 4, 2009, pp. 316-337.

[19] Siegel, S. & Castellan, N.J., Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill, Inc, 1988.

[20] Stevenson, B. & Wolfers, J., "Economic growth and subjective well-being: Reassessing the Easterlin paradox", *Brookings Papers on Economic Activity*, Spring, 2008, pp. 1-102.

[21] Thomas, R. & Millar, M., "The Effects of Material and Experiential Discretionary Purchases on Consumer Happiness: Moderators and Mediators", *Journal of Psychology*, vol. 147, n. 4, 2013, pp. 345-356.

[22] Uy, M.A.; Foo, M-D & Song, Z., "Joint effects of Prior Start-up Experience and Coping Strategies on Entrepreneurs' Psychological Well-being", *Journal of Business Venturing*, vol. 28, n. 5, 2013, pp. 583-597.

[23] Van Praag, B.M.S., "Well-being Inequality and reference groups. An agenda for new research", *Journal of Economic Inequality*, n. 9, 2011, pp. 111-127.

José Manuel Saiz-Alvarez is Academic Director of BA Doctoral Studies, Nebrija University, Faculty of Social Sciences, 28240 Hoyo de Manzanares, Madrid, Spain. He is also Visiting Professor at Catholic University of Santiago de Guayaquil (Ecuador), TEC de Monterrey, (México & Peru), and EAN University (Colombia). Professor Saiz-Alvarez has a doctorate in Economics and Business Administration Sciences from the Autonomous University of Madrid, and a doctorate in Sociology from the Pontifical University of Salamanca, Spain. He is President of the Institute for the Promotion of Economic Research and member of the Latin American Network of Family Businesses. He has been Honorary professor at the Autonomous University of Madrid. He is also member of the Editorial Board and referee of national and international journals.
Corresponding author: e-mail: jsaiz@nebrija.es.

Alicia Coduras Martínez is Director of the Chair of Entrepreneurship and an expert in quantitative methods. Since the year 1999 she has strengthen her focus on research related to entrepreneurship, heading the Global Entrepreneurship Monitor project in Spain until the year 2012. Now, as member of the GEM Coordination Team, she coordinates and manages the Expert's Survey (NES) at the international level. Professor Coduras holds a doctorate in political sciences from the University Pompeu Fabra and a degree in economic and business sciences from the University of Barcelona. She publishes regularly in academic reviews and develops different applied reports on diverse topics related to entrepreneurship for the Spanish governmental institutions.

Carlos Cuervo-Arango Martínez is General Director of BA & Tourism Doctoral Studies, Nebrija University, Faculty of Social Sciences, 28240 Hoyo de Manzanares, Madrid, Spain. PhD in Economics from the University of Minnesota (USA). Bachelor of Economics from the Complutense University of Madrid. Previously, he served as Executive Vice President of Banco Finantia Sofinloc (2002-2010) and Chief of Benito and Monjardín, Securities and Exchange Society (1995-2002) Director. Between 1988 and 1995 was Banesto Financial Officer, with responsibility for the areas of treasury, capital markets and financial planning and Bandesco President of Asset Management. Previously, he served as Deputy Director General of Monetary Management (1986-88), Research Fellow at the Foundation for Applied Economics Studies (1984-1986) and Consultant Training Centre of the Bank of Spain (1983-1984). He has published two books on Introduction to Spanish Economy and Financial System, as well as numerous articles on issues of monetary policy, banking regulation and structure of the financial system.

Intelligent Position Aware Mobile Services for Seamless and Non-Intrusive Clocking-in

Sergio Rios-Aguilar,

Pontifical University of Salamanca en Madrid, Spain

***Abstract* – This paper analyzes the viability of the use of employees smartphones as a valid tool for companies in order to conduct presence control. A Mobile Location Aware Information System is also proposed for a non intrusive Presence Control using exclusively terminal-based reactive location technologies, meeting cost minimization, and universal access criteria. The focus is providing trust to the employees, so that they feel safe and in control of when the location data is gathered while satisfying the control needs of the employer. LAMS platform is a state-of-the-art framework for synchronous mobile location-aware content personalization, using A-GPS terminal-based/network assisted mobile positioning techniques and UAProf data processing at the origin server.**

***Keywords* - Clocking in, Presence Control, Terminal Based Positioning, Network Assisted Positioning, A-GPS**

I. INTRODUCTION

There is enough evidence suggesting that use of Information Technology can play a quite important role on the growth of small and medium size businesses. In this line of thought, IT can be employed to bring about increased competitiveness if it enables businesses to create new jobs, increase productivity and sales achieving new levels of administrative efficiencies. It goes without saying that these outcomes can be achieved through measurable improvements of key performance indicators [1].

Tardiness, or –worst case- absenteeism, are two of the most persistent obstacles to productivity, profitability and competitiveness. They cause overtime, late deliveries, dissatisfied customers and a decline in employee morale amongst workers who are expected to cover for an absent or late coming employee. Early recognition and rapid intervention are key when it comes to managing absence in the workplace and can prevent absenteeism from becoming a long-term problem. [2] [18].

As for performance measurement, companies should collect and systematize all the information available so that they signal or allow the execution of their strategy to be successful in their business and remain in business. Once a company acknowledges tardiness and absenteeism as key problems, it should try to collect information detecting such behaviors.

Workplace surveillance and business organizations go hand in hand, and that employee monitoring is nothing new. The implication is that surveillance at work is, first, a necessity, and second, a normal, taken-for-granted element of working life. Employees expect to have their performance reviewed, objectives set, and information gathered on their activities and whereabouts – indeed, this is seen as good management practice. [18].

Controversies generally arise when employee monitoring goes beyond what is reasonable or necessary (i.e. when employers use what employees perceive as intrusive monitoring, gathering precise information as to how employees use their time) or when the application of monitoring negatively affects existing levels of autonomy and trust. [18][19]

So, when employers intend to control tardiness of absenteeism, they generally make use of clock-in systems. But they prove to be ineffective when applied to remote workers (ie. Sales force, field force…etc). [20]

Here we will be presenting LAMS platform, a Location Aware Mobile System for non intrusive clocking in. Field force employees make use of their own (or company supplied) smartphone to clock-in, and they are in full control of the positioning process and can trust that the system only gathers location data when they allow it to do so.

II. LOCATION TECHNOLOGIES AND LAMS OVERVIEW

One of today's key technologies related to advanced mobile services development is the physical positioning of mobile terminals requesting services. Moreover, the use of terminal based and network assisted solutions in order to achieve higher precision and lower acquisition times is the current direction of the state of the art research being conducted on this field.

In this paper, at first we will be presenting the fundamentals of Location Services and current location technologies, and we will go on with a description of the proposed platform designed to assist the employee clocking in process, with full HTML5/xHTMLMP-compliant contents personalization, together with integration services for spatial analysis, in terminal based and network assisted mobile services.

This implementation currently uses W3C Location API code within the terminal, in order to obtain the local position measurement from the embedded positioning device and the A-GPS data from a SUPL-enabled network. [14]. This information is then feeded to the JavaEE core of the platform, which build XML-based spatial queries and forwards them to the available GNSS servers, processes their XML responses (GML, PoIX) through XSL transformations to the final HTML5/JQueryMobile contents delivered to the requesting client terminal.[1],[2]

DOI: 10.9781/ijimai.2014.256

III. POSITIONING TECHNOLOGIES

Currently available technologies for physical determination of the position of a mobile terminal fall into two broad areas: terminal based technologies and network based ones.

In the former case, the positioning intelligence resides in the mobile terminal or in its SIM/USIM card. Within these technologies we have those based/dependant on GNSS systems (Galileo, GPS, Glonass), those which use the mobile network operators (MNOs) infrastructure (i.e E-OTD, Enhanced Observed Time Difference) and finally, those hybrid solutions which constitute the main focus in this paper: terminal based and network assisted positioning, currently represented by A-GPS.[9],[11]

In the latter case, network bases solutions don't require the integration of the positioning intelligence within the mobile handset. So, this kind of positioning services can be provided to all existing handsets with no distinction, as there are no change sin hardware required. The tradeoff is the relative lack of precision, comparing to the aforementioned terminal based solutions. Representative technologies in this area are CGI+TA (Celle Global Identity + Time Advance) and UL-TDoA (Uplink Time Difference of Arrival) [5][7]

A. A-GPS and Enhanced A-GPS

The Assisted GPS technology appeared recently and represents a key turning point. The technology enables a powerful hybridization between a worldwide location means – GPS– and a mass-market communications means – GSM/UMTS.

Moreover, Assisted GPS comes in handy mixing the best of the two worlds, since it compensates for the major faults of GPS and GSM/UMTS: a purely network- based technology does not provide sufficient accuracy (80 meters at best), and pure GPS solutions suffer from long delays before position delivery (typically several minutes). The principle of Assisted GPS consists of coupling satellite positioning and communication networks, sending assistance data to the receiver integrated in the handset to improve its performance [7][10]

Compared with standard GPS, Assisted GPS offers (1) very short latency to get a position, by sending satellite data much more rapidly, than GPS itself; (2) Hence, very low power consumption; and (3). Increased sensitivity, therefore increased availability of the location service, particularly in dense urban area and indoor environments.

Enhanced A-GPS is an improvement of the classic A-GPS positioning with new satellite technologies: EGNOS (European Geostationary Navigation Overlay Service), and Galileo. The Enhanced A-GPS provides EGNOS-based assistance data to GPS-enabled mobile phones, via GSM/GPRS or UMTS networks. This is obtained by incorporating an EGNOS/GPS reference receiver in the Enhanced A-GPS server, which receives the EGNOS correction messages.[11]

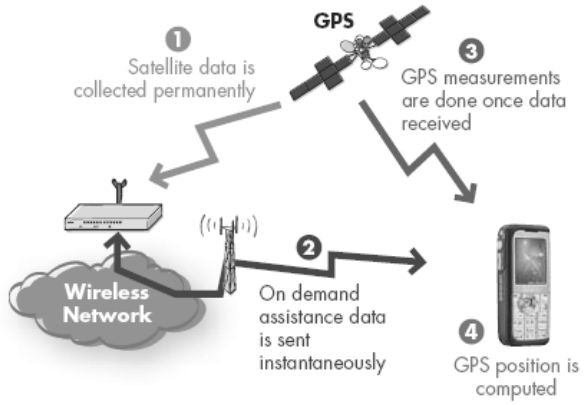


Fig 1. A-GPS Operation Flow

EGNOS is a system consisting of three geo-stationary satellites and 34 ground stations in Europe, the Americas and Asia. EGNOS improves GPS & GLONASS position accuracy and availability by 'adding' measurements from its three geo-stationary satellites locally visible in Europe to the Galileo, GPS & GLONASS, and by providing ionosphere, orbitography and clock corrections.

IV. LOCATION AWARE MOBILE SERVICES (LAMS) PLATFORM

The LAMS Platform has been developed as a proof-of-concept of the proposed service architecture, acting as an universal LCS (LoCation Services) client.(See Fig.3)

Once the platform gets the positioning info obtained using local AGPS data through W3C Location API, sends it to the server , which in turn, trans codifies this information to XML, using an specific XML application (LAMSX). With this information, the platform sends a spatial query to a GIS server in order to obtain reverse geocoding data which carries the desired Points of Interest (POIs) also in native XML format. [4],[5],[6]

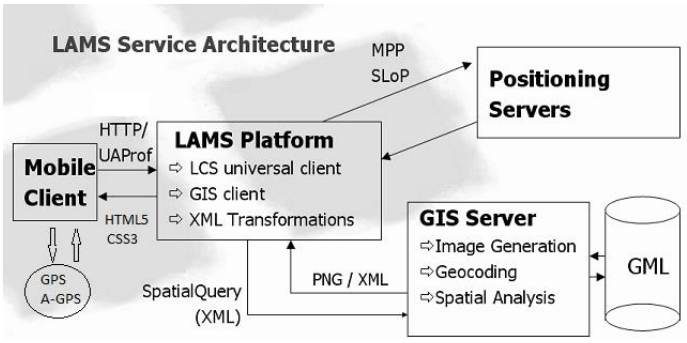


Fig 2. LAMS Platform Service Architecture

At this point, the platforms enters a dynamic content generation phase, personalizing the overall response to the make, model and capabilities of the requesting handset, so that the employee gets the best experience in the clocking in process, no matter which phone he or her has been using to complete the task [3]

Finally, the outcome of this content generation phase is the end HTML5 / JQuery Mobile or xHTML/MP page targeted to the handset.

V. CONCLUSIONS

In this paper has been presented the practical use of Location Services, applied to non intrusive field force clocking in, allowing the employee for the control of their position data gathering times.

This knowledge can be used along with the also described UAProf framework to achieve full personalization of the contents delivered to the end client. In this case, we are able to dynamically generate location-dependent, device-dependent and user preferences-aware contents.

Also, this paper demonstrates the feasibility of a time-constraint service provision when using terminal based technologies with the most up to date enablers, such as W3C Location API, OMA SUPL and A-GPS at application level.

REFERENCES

[1] De Assis Lahoz, M., & Camarotto, J. A. (2012). Performance indicators of work activity. Work (Reading, Mass.), 41 Suppl 1, 524–31. doi:10.3233/WOR-2012-0207-524

[2] Ball, K. (2010). Workplace surveillance: an overview. Labor History, 51(1), 87–106. doi:10.1080/00236561003654776

[3] Franlin Reynolds, Johan Hjelm et al. Composite Capability/Preference Profiles (CC/PP): A user side for content negotiation. World Wide Web Consortium http://www.w3.org/TR/NOTE-CCPP

[4] Ríos Aguilar, S. Generación dinámica de contenidos WAP para terminales móviles, Libro de Ponencias del Congreso "Mundo Internet 2000", February 2000

[5] Ríos Aguilar, S. Position-aware WAP Contents Personalization. Wrox Wireless Developer Conference, Amsterdam, July 2000

[6] Ríos Aguilar, S. Interfacing to a Mobile Positioning Center. Wireless Developer Network (www.wirelessdevnet.com), September 2004

[7] Schiller Jochen, Voisard Agnes. Location Based Services. Elsevier-Morgan Kaufman 2004

[8] Spinney Jonathan, Wireless Location Uses in the User Plane and Control Plane. http://www.lbszone.com/content/view/148/45 Visited in May 2007

[9] Stojmenovic I. Handbook of Wireless Networks and Mobile Computing. John Wiley & Sons, New York, 2004

[10] Katz, J.E. , Aakhus M. (eds). Perpetual Contact. Mobile Communications, Private Talk, Public Performance. Cambridge University Press, Cambridge, UK 2005

[11] AGILE project. http://www.galileo-in-lbs.com/index.php?id =398. May 2007

[12] OMA-TS-MLP, "Mobile Location Protocol", Open Mobile Alliance, http://www.openmobilealliance.org/

[13] OMA-RD-SUPL-V2_0, "Secure User Plane Requirements v2.0",Open Mobile Alliance, http://www.openmobilealliance. org/

[14] JSR-179 Location API for JavaME. Sun Microsystems Java Community Process http://jcp.org/aboutJava/communityprocess/final/jsr179/index.html

[15] Wu Shiow-Yang, Wu Kun-Ta, Effective Location Based Services with Dynamic Data Management in Mobile Environments, Wireless Networks 12, 369-381. Springer Science, 2006

[16] Küpper Axel, Treu Georg, Efficient Proximity and Separation Detection among Mobile Targets for Supporting Location-based Community Services. Mobile Computing and Communications Review, Vol 10, No. 3. 2008

[17] Xu Toby, Cai Ying. Location Anonymity in Continuous Location-based Services. Proceedings of the 15th International Symposium on Advances in Geographic Information Systems. ACM GIS 2007

[18] Ball, K. (2005). Organization, Surveillance and the Body: Towards a Politics of Resistance. Organization, 12(1), 89–108. doi:10.1177/1350508405048578

[19] Axtell, C. (2011). The Well-being of the Mobile Workforce. iPass 2011 Report.

[20] M. López-Fernández et al.. "Control of attendance applied in higher education through mobile NFC technologies". Expert Systems With Applications Vol 40, Issue 11, September 2013

Sergio Ríos Aguilar holds a PhD in Economics from University of Granada. He also holds a MSc in Telecommunications from King Juan Carlos University (URJC) and Computer Sciences degree from UOC. He currently works at UPSAM teaching Mobile Development and is conducting research on advanced mobile enterprise information systems

Bring research out of the lab, into real-life scenarios

- BIG DATA
- CLOUD COMPUTING
- COMPUTER VISION
- ROBOTICS
- INNOVATIONS IN ARTIFICIAL INTELLIGENCE
- INTELLIGENT SECURITY SYSTEMS
- INTELLIGENT MOBILE TECHNOLOGY
- SEMANTIC WEB
- IJIMAI.ORG

YOUR BEST RESEARCH PARTNER

imai solutions

Advanced Research Systems

http://www.imai-solutions.com

•Addresses your research and development issues in a way which develops your reputation and sharpens your competitive edge

•Our own global experience enables us to understand your business needs and empowers us to ask the right questions that lead to meaningful results.

•Opens up new and surprising creative avenues that can nourish and energise your organisation's design thinking

•Enables you to build a relationship with a extended network of researchers.

Paseo de la Castellana 95, Fl 15 - TorreEuropa. 28046 Madrid , Spain. T:+34914187921/F:+34914186999. info@imai-solutions.com

www.elasticbox.com
info@elasticbox.com
blog.elasticbox.com
Twitter: elasticbox

elasticbox

Deploy applications, not servers.

Challenges In Cloud Computing

Automation

- How long to deploy an application?
- What version do I use?
- How do I upgrade applications?

Portability

- How do I change providers?
- What is being used?
- How much does it cost?

Auto-Scaling

- Can my application auto-scale?
- How do I configure auto-scaling?

Disaster Recovery Planning

- Can my application tolerare faults?
- How do I recover my system?

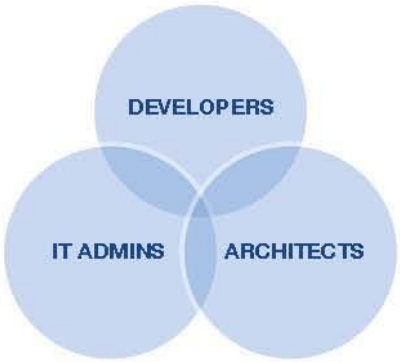
ElasticBox Solution

Automation

- ✔ Automatic Deployments
- ✔ Automatic Configuration
- ✔ Disaster Recovery

Runtime Environment

- ✔ Application Scaling
- ✔ Fault Tolerance
- ✔ Resource Clean-up
- ✔ Replication



Framework Design

- ✔ Architecture Policies
- ✔ Versioning
- ✔ Platform Management

Infrastructure Control

- ✔ Cost Analysis
- ✔ Policy Management
- ✔ Traceability

amazon web services™ · Windows Azure · hp · rackspace · openstack · EUCALYPTUS