Using Social Networks based on Data Mining to Promote New Technologies for Women from Emerging Countries

Syrine Ben Meskina 1, Valérie Monfort1,2 and Achraf Ben Miled 1

1 SOIE, 41, Rue de la Liberté, Cité Bouchoucha 2000, Le Bardo, Tunis, Tunisia
2 Université de Paris1, Panthéon Sorbonne, France

Abstract. Nowadays Social Network Websites become increasingly numerous and millions of people are registered. They become famous and useful in several domains, in particular for sales and marketing activities all over the world. Their success is gradually growing in industrialized countries. In our paper, we firstly show how this technology is used to promote and enlarge sales networks. Secondly, we investigate the ontologies dedicated to represent the Social Network Semantics and their limits. We propose to use this approach to a research project which aims at addressing women from both rural and urban areas of emerging low income countries, who are individually or within group involved with sustainable hand craft and/or fair trade business leading to a varying degree of economic mobility. We finally conclude by proposing a solution that aims to promote Sales Social Networks based on Data Mining techniques.

1 Introduction

Nowadays, it seems easy and usual to utilize new embedded applications on PDA and Social networks. Past months showed impacts of these technologies during social movements. Moreover, economical analysis showed companies aim to keep their clients and to promote one to one relationship to increase their sells and their business. They use social networks, Web sites, mails, and SMS, to promote and improve their products.

However, what is obvious for industrialized countries is not so easy for emergent countries. Costs of the devices such as laptops, PDA, software developments to develop Web sites and simple mobile phones are heavy for emerging countries. Moreover, technologies are not disseminated and technical infrastructures as networks and relay antennas do not cover the whole country. Sometimes it is not possible for electronic transaction to pay in Euro or Dollars, or by credit cards.

Based on a research project, we address women of emergent low incomes countries. These women are involved and are developing sustainable, hand craft and/or fair trade business. The project is structured into several main steps such as:
- Elicitation: understanding the manner women use new technologies and social impacts,
- Defining and improving behavioral patterns to adopt new technologies
- Storing and analyzing behaviors
Proposing new behavior, technical solutions, and training

These steps are the cornerstone of a process to improve iteratively. The aims are to improve business of these women by offering them true technological means adapted to their habits and to the technical and legal contexts of their countries. Moreover, it is a way to think about how to propose to the greatest part of the humanity new cheaper technologies. So, to develop a specific affordable business for emerging countries. We focus on the ontologies used to characterize users in emerging countries. Moreover we propose a manner to analyze information to understand how new technologies are used by women in emerging countries and if they have an added value for business. We aim to measure the impacts and the benefits of the new technologies and the usage of the propositions we can do to improve business.

Thus, we briefly describe the project, secondly, we introduce and define the basic relative concepts; Social Networks and the ontologies dedicated to them as well as the relative social and semantic problems. In the fourth section, we will briefly cite the main limitations of the evoked ontologies and then mention, as a solution, some techniques of ontologies composition. Section 6 shows related works. Then, in the remaining sections, we will present our proposed work. Section 7 concludes the paper and outlines the future work.

2 Project Presentation

Present status of Web 2.0 is far advanced than that of the basic DiNuccian ideology in this concern, and the modern Web 2.0 is smartly admixed with the facilitation of participatory information sharing, interoperability and user centric design where the Social Networks are considered as the cornerstone of this system. So, the development of Business Intelligence techniques with Web 2.0 and Internet offer an access to less expensive and/or free resources and provide opportunities to economic and social actors. The essential of these opportunities and resources aims to transform time into information to enable a better decision making. Enterprises and e-commerce are the most impacted areas by this evolution.

However, if this artifact seems coherent and promising for companies, its implementation remains limited because it is faced to several constraints such as the unavailability of fitted infrastructure as well on the technological aspect, a range of restrictions by laws and hindrances from a degree of regional / global politics. These conditions are applied in developing countries; the situation remains so complex that researchers use the expression: new technologies excluded or digital divide, by showing the possible relations with social and economic or banking exclusion.

To face to this new kind of gap between societies, several research works have been realized to find the adequation between new technologies and usage capacities of the wider part of population targeted. Most of the peoples of these countries have weak access to the benefits of new technologies and e-commerce.

During last past years, several works have been aiming to offer technological products and services in adequacy with means, needs and social habits of populations in emerging countries. These domains concern with well education, health, agriculture, telecommunications, e-commerce, etc. Companies are also trying to enter these
markets by using sells strategies along with market structure and its needs (such as rural phone).

One of the starting points of these works is the incorporation of the concept of “Base of the Pyramid”. This concept is based on a simple idea, i.e. 4 milliard people who live with less than few dollars per day represent a set of opportunities in terms of demand of goods and services. These set of opportunities is linked to the “non inclusive” strategies deployed by economic actors, private or public, particularly for emerging countries.

Till the publication of first research works about the concept, and critics, several projects and programs were launched. Some of these works show the importance of the new technologies in the socio-economic development of this population. The new technologies improve the socio economic situation of this population (Base of Pyramid) by bringing them information and expertise for consumers and managers. However the results of these works remain limited and a global approach should allow combining a local understanding of habits and needs of this population with a technical expertise of the product or service, and a sectoral expertise. The ‘applied’ part of the global technological innovations in this particular concern is required to be flexible enough that it can be necessarily molded as per the physico-socio-politico-economic characteristics of regional domain of application.

Firstly, this project is based on a global approach based on synergies between specialists in different domains in particular: sociology, economics, computer science, and marketing. It aims to study the impacts of new technologies on populations of both countries, Algeria and Tunisia. Focus to be given on two actors categories: very small companies and members of formal or non formal social networks. Principal aim is to address these two categories and to study the impact of technologies on business activities proposing new habits to use new technologies. This approach is fully iterative and incremental. Knowing current habits and usage pattern by auditing and eliciting users allows noticing lacks and proposing new approaches. The different proposals and its effects are audited and improved.

Secondly, the result of this study can imply the requirements of new kinds of affordable technologies fitted to emerging countries. Moreover, these new technologies have to be adapted to the profile of the users, for instance, the users can be analphabets and the media have to support text to speech services. The definition of new paradigms for a great part of humanity has the consequences to create start up to commercialize new packaged products and allied services. On top of all, generative programming technologies are an issue to automatically define and generate applications without requiring costly investments in developing e-commerce or M-commerce applications.

Thirdly, this project requires to be based on a performing technical infrastructure to analyze information usage and to help users to learn new technologies usage. Moreover, this infrastructure is based on: Web services providing service based approach and interoperability between applications, social networks, mobility, etc. So, proposal may be given to develop and stabilize this architecture to support marketing and social research works, and to promote a true course of actions to help and train women to new technologies and future fitted measures for improving business.

The proposed research will involve triangulation among a variety of different sources of data. Hence, we intend to conduct both formal and informal on- and off-site interviews with actors. We will also analyze archival materials such as internal
documents as well as articles and evaluate existing case studies and other relevant literatures.

Case studies may be needed to deduce a common knowledge and identify stages and obstacles impeding progress from one stage to the next, as well as characteristics of projects that are believed to successfully navigate through the processes. Depending on the measured intensity of factors, it will be possible to classify different technological orientations and to predict on which development stage a sector or a country is situated.

Pragmatically, the project will be divided into four main steps as follows:

- Firstly, the behavior of Tunisian and Algerian professionals is observed and the way to use of new technologies as well as social and commercial spheres will be studied. These behaviors are based on formal and informal knowledge to collect and analyze qualitative and quantitative data with rules. These techniques were already used to collect qualitative and quantitative data. Knowledge on Algerian and Tunisian professionals will be modeled, structured and validated according to business experts, computer scientists, marketing experts and sociologists. Sociologists, marketing experts and computer scientists provide a behavioral and habits patterns repository proposing and improving fitted solutions to facilitate the integration of new technologies and social networks during commercial transactions initiated by professionals. These knowledge are parameterized, aggregative, changeable, modifiable and flexible, according to functional, non functional, and technical needs, but also according to ambient context, economical and social environment.

- Secondly, the study will link the current behavior of professionals faced to new technologies and social networks to offer solutions or improvements.

- The third step proposes to dynamically updated users’ profiles and behavioral patterns through the data from data mining analysis, data warehouse, marketing analysis and sociometry. This approach is iterative and incremental according to the manner users appropriate new technologies and social networks usage. New manners are proposed with feedbacks. In the same time innovative approaches and/or technologies is to be framed.

- The fourth step proposes to use innovative solution defined in the third step, to check the way it is appropriated and to train people to improve their business.

3 Basic Concept

3.1 Social Networks

A Social Network is a dynamic social structure composed of nodes and canals or arcs such that the firsts represent individuals or organizations and the lasts represent their relationship.

According to [11], a Social Network is, generally, defined as a relationship set of a specific type (for example: collaboration, support, control ...) between a set of actors. This definition calls the Barnes’s notion of mesh in relationships as well as the traffic notion to transmit data between individuals over the mesh.

In Sociology, the Social Networking definition is different from Social Network
one. In fact, it references the employed means to inter-connect individuals. Social Networking refers to multimedia social structure which essentially facilitates communication between individuals, groups and organizations. Since the apparition of the Web 2.0 having a user-centered design, the Internet users become actors and the networking concept has been extended to sites building social networks on the Web.

Social Networks are based on connections between virtual identities which are created on the Net and which may be veritable or not. So, the profiling concept was included by the majority of the Social Network definitions. The main characteristics of the classical ones are turning around that profiling notion and are summarized on; create profile which represents a virtual identity necessary for connecting to some members of the Social Network, look for friends with whom the relationship is desired and finally draw links with other profiles which is the aim of the Social Network. The most commonly recognized social networking Web sites include Facebook (specially in Tunisia), LinkedIn and MySpace... They are used for professional, cultural, sport activities and specially to do marketing and for sales. For the last usage domain, we call them Sales Social Networks.

In addition, to social Network sites, there are other ways to accomplish social media marketing through blogging, instant messaging, widgets ... which become increasingly popular internationally. Using Social Networking as a form of social media marketing is the new fad.

Facebook, blog platforms and other Social Networks have proven to be essential for sales people as well (For example: in Tunisia, all sales persons or people have at least one page or profile on the social Web networking. It allows them sharing ideas and digital photos with their friends and publishing real as well as virtual events to add value...). Social Networks present the fastest, cheapest and easiest way to know more new customers, new ideas improving and enlarging their business. Technically, Social Networks are based on specific meta models called ontologies.

### 3.2 Social Networks and Ontologies

Web becomes “social”. Hence, it is necessary to use a common model to represent the semantic of the majority of these Websites. Ontologies, mainly are defined by classes and properties, and provide common semantics for resources on the Web semantic according to [6]; [9]; [1];[4].

To represent social data, some common representation models are dedicated to the social Web semantic adopting the idea that semantics can help social Websites and vice versa. In fact, social media sites can interoperate by calling common semantics, defined using agreed-upon semantic formats, to describe people, content objects and the connections that link them all together. Current data on the Web own different formats such as: relational Data Base, XML, CSV... Hence, the integration and the link of these data become complex and infeasible. For instance, RDF presents a standardized way to publish data on the Web thanks to its triplets represented by labeled graphs.

(Berners-Lee, 1998) has introduced the idea of Linked Data which are based on URIs for naming, identifying resources, providing information about them and including links to other ones in the aim to discovering more things. Within the framework of LOD Project (Linked Open Data), the existing datasets are translated into RDF and
linked to each other (for example DBpedia and Geonames, Freebase, BBCprograms …).

As users may dispose of different accounts on the same or different social Web-sites on the Web, it is interesting to establish links between them.

FOAF (Friend of a Friend) presents an ontology designed by Tim Berners-Lee for describing people and the existing relationships between them using the RDF standard vocabulary. It mainly investigates the identity notion, personal profiles and social networks and can be integrated with other Web Service vocabularies. It aims to repre-sent the concept of agent and the different related subclasses (persons, agent groups, organizations as well as some properties related to these concepts such as names, knowledge, membership…) based on the management of the distributed identities over the Web. Consequently, it inter-relates different processes as interlinking identities and social networks to associate individuals to all their online accounts over the social Web by exporting data from different sources. It is used for FriendFeed, LiveJournal, MyBlogLog…

SIOC (Semantically-Interlinked Online Communities) ontology is defined to de-scribe the online activities of communities on the Web such as: blogs, forums, wikis, etc. In fact, it fully describes the content and the structure of the social Website by representing the socio-cultural data; communities and their activities, produced doc-uments as well as their organizations and inter-connection. It creates a kind of bridge to link several Web tools to collect the different elements of communities from separate places on the Web. It is composed of a core of eleven classes concerning the social aspect (user accounts) as well as the structural aspects (contents and containers) and properties describing their relations. It includes two main modules such as: Ser-vices and Types concerning the social Web content. SIOC uses the RDF vocabulary to describe the previous referenced elements. It also uses objects defined by other ontologies such as: FOAF, SKOS, DublinCore, RSS etc. The goal of SIOC ontology is to address interoperability issues on the social Web when it is combined to FOAF (Bojares and al., 2008).

The tagging concept consists of associating some information to resource. It al-lows sharing, exchanging and reusing of the data that was tagged over the social Web. The TagOntology proposed by (NewMan, 2005) is based on the Gruber’s tag which is composed of three parts such as: i) the tag representing the word or the sentence recognized by humans and machines, ii) the annotated resource representing the part to be tagged and that must be identified using an URI or another similar labeling service, and finally iii) the tagger representing the user which indicates the tag. Tags should be named to differentiate them to create them such as resources which can be uniquely identified. The contribution of the TagOntology consists on adding a tagging visibility property.

SCOT (Social Semantic Cloud of Tag) ontology is designed for representing the structure and the semantic of a tag set. More exactly, this standard model describes tag clouds (tags and co-occurrence), allows the portability to move from its own tag cloud to another one. It, also, provides the possibility of developing the global share (sharing tag clouds between services and between users). An important contribution consists on supporting the social interoperability between heterogeneous sources.

MOAT (Meaning of a Tag) ontology is a standardized model designed to define tags meanings. It provides a Web Service structure which supports the publishing of semantically enriched content (free tagging). In fact, it allows users associating mean-
ings to tags via URIs on the social Web. It is a collaborative approach as it shares meanings in a community and brings the tagged content to the linked data Web.

CommonTag is close to MOAT and allows associating tags to meaningful resources. NiceTag Ontology appeared for which tagging meets speech act theory and investigates on the link between a tag and a tagged item. It is, also, possible to go in the opposite way by extracting ontologies from tags thanks to FolksOntology which is a semi-assisted of relationships between tags.

It is also interesting to localize entities. In fact, it is relevant to geographically identify the social network as well as its members and partners. It links, for example, two instances to the same subject using the GeoVocabulary. It is also possible to use Geonames to provide geographical data.

Moreover, geospatial ontologies localize data according to space and time.

3.3 Problems Met

There are many problems relative to Sales Social Networks. They can be categorized on two groups.

The first type is social. In fact, to preserve the profitability of its business, it is important to point that the adoption mobile technologies are primordial (being always connected and reachable using the smart phones and laptops, to be up to date...)

The emergence of smart phones and laptops with small screens seems to be more attractive especially for emerging markets. But the prohibitive costs of high-end smart phones, high-speed Internet, additional memory and new technologies, in general, decelerate the development of mobile social networking. On the other hand, emerging markets are behind on mature markets of payment, remote buying, using credit cards, no delivery... this is due to the non-adaptation of the infrastructure.

Moreover, for instance, there is a very low Internet penetration in many parts of the world especially for emerging countries seen its high cost. It is, also, useful to take into account that social conditions and NTIC usage vary widely between emerging countries capital and regions.

The second type concerns technical problems. According to the Semantic Web Layer Cake designed by (Berners-Lee, 2001), each layer is composed of one or more languages and standards dedicated to the ontology manipulation. We will present their most important limitations. RDF language was appeared in 1999 for describing knowledge using triplet of the form (subject, predicate, object). It provides only very simple and elementary RDF descriptions and dispose of a reduced number of predefined constructs.

RDFS language extends RDF and resolves partially the problems relative to manipulating non accurate data. Using RDF and RDFS present numerous advantages such as associating namespaces to schemas, supporting multiple heritages as well as obtaining some descriptions for which several classes are instantiated at a time.

3.4 Data Warehouse and Data Mining

Social Networks Websites dispose of huge databases leading to data storage problems. Even if such Web sites can be represented through ontologies. These data may
be vital for marketers to be analyzed to promote sales and marketing activities. Thus, the corresponding studies start from the past data to the actual data going to preview the future in order to describe the customer behavior, from which comes the necessity of storing all information about customer.

It is important to recall that user data are usually confidential, thus, marketers aims to consider even the smallest details about customers as they may be useful which increases the amount of data.

Data warehouse and data mining techniques can be helpful to identify the pertinent data to be considered and to identify and extract the most important rules and properties in the aim to refine analyses.

4 Towards an Ontology based Solution

4.1 Current Ontologies Selection and Limitations

In the Semantic Web, knowledge and data are generally represented through ontologies supported by the tagging notion (section 3.2) for the social Semantic Web.

In this issue, several standard ontologies were developed during the last decade. The well known and most used ones were exposed in section 2.2. However, the majority of them, when they are used separately, seem to be limited and insufficient to overcome some problems mainly summarized on distributed sources, heterogeneous data, tags ambiguity, semantic links between tags... seen that no ontology works under all conditions. Hence, it is interesting to consider a set of them at a time.

The combination of ontologies is formed using composition techniques which will be exposed on the following section.

One such combination example can be compound of TagOntology, SCOT, SIOC and MOAT. This model supports tags, simple and tripartite tagging, agent modeling, definition of cloud of tags and associating meaning to tags at a time.

4.2 Ontologies Composition

The representation of some systems requires the intervention of more than one ontologies which can be heterogeneous and distributed. So, the cooperation of ontologies is recommended as they may solve such problems (Euzenat and al., 2004).

The approaches dedicated in this purpose use the mapping and alignment techniques. They, respectively, consist of identifying similar entities from different ontologies and establishing mapping links between distinct representations.

The manipulation of ontologies includes comparison and inter-operation.

The first axis, also called “alignment” processes by mapping equivalent entities from different ontologies using the adequate metric (Johnson and al., 2006).

The second axis describes, in general, the inter-operation between ontologies by both preserving and extending them. It is a sort of combining heterogeneous or distributed ones based on mapping schemas without being modified (copy, alignment...). Generally, it is useful to combine the ontologies that describe complementary domains. The fusion technique allows the creation of only one global and coher-
ent ontology. Thus, the resulting ontology must unify knowledge described by the yet existing ones (importing the whole original in another one).

The integration technique provides an ontology created by considering only some proportions from the other original ones (it is different from the complete fusion).

The coordination technique consists of using, at same time, the knowledge described by only some parts of the original ontologies in an independent manner.

5 Towards a Data Mining based Solution

The problems of promoting Sales Social Networks depend on the collected data about customers. The data may be analyzed for useful correlations between interests and also between interests and demographic categories.

The reasoning activities over social networks are usually based on the proposed ontologies to represent these Web sites (for example each Website is represented by ontology). Then, to model the users of Social Network Websites, we can collect all information concerning them from other social sites by interlinking several Social Websites in order to better describe the different customer classifications.

This process requires the use of ontology composition techniques. We can use the combination model evoked in section 4.1 to represent user profiles. In a second step, it is interesting to look for forming a global ontology on which we can effectuate some reasoning and inferences in order to preview new marketing activities based on NTIC use.

We recall that (Klein, 2001) defined mapping by finding equivalent elements according to a similarity measure even if they come from different data representation (ontology, data base schema). We mention also that Social Networks have huge data bases containing information about their users. It is possible to proceed by mapping the real stored data bases and the proposed ontology to resolve our problematic.

In the other hand, we plan to implement some data mining techniques over the considered data bases which consist of identifying the pertinent information by using data reduction in a first step. And, then in a second step, effectuate a mapping process between the ontology and the kept data.

6 Related Works

In the literature review, we focused on several research works as our project handle different domains; marketing, Semantic Web, ontology manipulation, Web Mining...

Bouzeghoub and Kostodinov [3] studied a personalization of profiles in order to facilitate the need expression for a particular user or community and to obtain relevant information from an information system. They defined a set of criteria and preferences specific to each user or community of users and proposed a generic model of profile allowing classification tasks over profile’s contents. Slimani [16] defined some approaches for evaluating semantic associations on domain ontologies in order to discover heavily linked data collected from separated sources. Jarrar [10] presented and used ontology for customer complaint management which has been developed in
the CCFORM project. As customers can register all their complaint against any party about any problem, it can be useful to explore this data in the aim to promote Sales Social Networks. In order to facilitate the re-use of ontologies, Maedech [15] proposed a ontology learning framework enhancing typical ontologies engineering environments for importing, extracting, pruning and evaluating ontologies. Berendt et al. [2] discussed the combination of Semantic Web and Data Mining, the use of Web structures to do Web Mining and vice versa as well as the resulting profitability of integrating closer data. Stumme et al. [17] proposed an automated schemes for learning the relevant information to automatically enhance several resources as it becomes impossible to do it manually seen the enormous size of data on the Web. (Gruber, 2007) discussed the combination of the best ideas from the Semantic Web and the Social Web which depends of individual user contributions. He proposed collective knowledge systems which unlock the “collective intelligence” of the Social Web with knowledge representation and reasoning techniques of the Semantic Web. (Li, 2003) presented an approach for automating the discovering of ontologies from data sets in the goal of building complete concept models for Web user information needs. The proposed approach overcomes the problem of obtaining information from user Web profiles. Li [14] evaluated the previous model to be able of capturing the evolving patterns to refine the discovered ontologies and assesses their relevance. Li [13] presented the relationship between Web mining and linked data such user profiles and proposed an abstract Web mining model for extracting approximate concepts hidden in user profiles on the Semantic Web based on ontologies. The developed approach includes an efficient filtering algorithm to filter out most non-relevant inputs.

We found that ontologies handling the context and the interest criteria can be useful to resolve our problematic as we look for adapting a user profile to the women previously described.

There exist some ontologies which investigate the context of Web pages represented by documents. The Foafnaut ontology, proposed as an extension of the FOAF ontology, describes one’s contacts details based on the associated context. That family of ontologies enables us to measure the relationship between persons by calculating scores using keywords and related contexts.

Contexts depend of a periods of time, location, documents and Web pages as well people’s preferences and interests. As contexts change interests may change. The Weighted Interests Vocabulary, specified by Bob Ferris in 2009, attributes weights to people interests in dynamic manner. In fact, it includes both of E-FOAF: interest vocabulary which extends the FOAF vocabularies, in 2010, to describe user interests and preferences (interested by, disinterested by as well as preference degrees…) and the Interest Mining Ontology.

The Geo ontology is also a FOAF extension proposed by Dan Brickley in 2009 to specify the coordinates; latitude, longitude and altitude of objects in the Web.

As research works treat separately each one of the domains on which we are focusing, it is interesting to work under the combination of all of them.

7 Conclusions

We propose the first step of a research project based on new technologies such as:
social networks, mobility, e-commerce. We offer a possibility to proposed fitted solution to communicate about business such as aircraft. As social networks are based on Web 2. And ontologies we proposed an overview of current existing ontologies based on social networks. We have to use and/or to modify existing ontologies or proposing new one. We also have to link these ontologies with data warehouse and data mining by proposing fitted algorithms to manage data and their instances. We aim to improve this approach and to propose concrete results after interviews, structuring and collecting data.

References


