Cyber-Physical Directory : A Dynamic Visualization of Social Media Data

Jean-Loup Lamothe, James She and Ming Cheung
HKUST-NIE Social Media Lab., Department of Electronic and Computer Engineering,
Hong Kong University of Science & Technology, Hong Kong, China
email: jlarlamothe@ust.hk, james.she@ust.hk, cpming@ust.hk

***Abstract*—Information about people, places and events are examples of social data. These social data, widely spread on the Internet, are also displayed by directories directly in the physical world, enhancing interactivity and guiding people at a specific location. However, they are usually large, and reading these directories is time-consuming because they are not personalized, with information unrelated to a person's needs. Moreover these social data are static and may be irrelevant for many readers. This paper presents a new cyber-physical system: the cyber-physical directory, which provides a user with a customised and dynamic visualization of social data. An algorithm, based on the similarity between people and social data, finds which data are relevant to a specific user and displays them by using tagcloud techniques. The system is successfully tested with a real dataset from Foursquare, and an implementation is presented.**

I. INTRODUCTION

Every technology is primarily used for technical purposes before turning to broader uses. The web 2.0, with websites, blogs and social networks has transformed the Internet from a simple communication platform to a huge online social place. Also, phones have become smart and able to communicate intelligently, increasing the interactivity among people. Moreover, since location technologies such as GPS are embedded into mobile devices, social interactivity has begun to be transferred from the Internet into the real world and the spreading of social interaction has increased dramatically. Though online recommendation systems can sort this big flow of online data and provide a user with personalized content, enhancing online interactivity, such recommendations do not exist for social data directly displayed in the physical world. Fig. 1 shows that a physical or digital directory provides much social information, such as events or places, to increase the interactivity at a specific location, but this large amount of information may be irrelevant and difficult to read for some people. Digital directories, like those presented in Fig. 1, are more convenient, eye-catching and attractive than physical directories. Some provide a search function that enables users to find specified topics related to their needs. Others, like those proposed by [1], integrate a touch screen that improves the user experience. The user may then access interesting information. However, they present some limitations. The increasing number of interactions for a user may decrease the number of people who are able to access the system, especially in crowded places. Moreover, though they present rich and well-structured information, the user still needs to manually control the system before obtaining the required information from the system.

Fig. 1: physical (a) and digital (b) directories

Today, a new generation of devices called cyber-physical systems are generating increasing interest. In the last few years, cyber-physical systems have become an important field of research [2]. A cyber-physical system is described by [3] as a system which actively engages with the real world. [4] refers to a system that is a combination of a computation system and elements in the physical world that can interact with humans. This kind of system is now applied in many areas such as home control, traffic control and health. Nonetheless, like the Internet or phones, cyber-physical systems can also be considered for social applications. Present in the physical world and able to interact with humans, they can be used to provide users with personalized content at a specific location.

Motivated by the above observations, the concept of the cyber-physical directory is proposed. The cyber-physical directory is an improvement of the current directories. It collects profile data from people around it through smartphones and changes the visualization of its social content according to people's interests. This system does not only display its social content but includes the cyber-physical social relationship between a user and its social content such that the visualization of the information can be adjusted dynamically. The proposed visualization, based on tagcloud (also referred to as wordcloud) techniques, gives greater prominence to information highly relevant to a specific user. A Jaccard-based similarity algorithm is created to find which information is relevant by analysing the profile of a user with the social content displayed by the system. Such a profile can be built from a social networks' Application Programming Interfaces (API) or provided directly by a user through a smartphone application. When a user goes closer to the proposed system, his or her profile is sent to the system through a smartphone. From these social data the system creates dynamically a personalised visualization of the information that enhances the information that may be relevant to the user.

This paper brings the idea of a new social cyber-physical system, proposes an algorithm to find relevant social information and deals with a new way of using tagclouds as an efficient and personalized visualization.

The rest of the paper is organized as follows: Section II describes the system architecture. Section III presents the datasets used in this paper, the similarity algorithm and the experimental results. Section IV concludes the paper and presents future works.

II. CYBER-PHYSICAL DIRECTORY

A. *Overview*

Three scenarios are considered to show how the cyber-physical directory works and displays social content. This social content is composed of social things such as events, places or people. Firstly, Fig. 2(a) considers a regular directory of people versus a cyber-physical directory of people. In this case, when a user goes closer to the cyber-physical directory, the visualization of the people changes by using tagcloud techniques and the size of each picture is adjusted dynamically according to the user's interest. This new visualization gives greater prominence to people who are likely to interest the user and changes dynamically depending on the user's closeness to the system. Secondly, a restaurants directory is presented in Fig. 2(b). Here, the proposed system displays the restaurants that may interest the user with a greater size, according to the user's dining preferences. Finally, this system is also convenient to display a list of events, as in Fig. 2(c). Although a regular and a cyber-physical directory show the same information, the visualization proposed by the cyber-physical directory is more efficient for a user seeking an interesting event, according to his or her cultural background. By changing the visualization of the information and by giving greater prominence to relevant information, the cyber-physical directory enables people to quickly find the information which most interests them.

Fig. 3 shows the system architecture of the proposed cyber-physical directory, which mainly involves: 1) a digital display (a screen for example) with a wireless access point; 2) a computational system embedded in the digital display and 3) smartphones. The wireless access point provides Wi-Fi access to enable the digital display, the computational system and smartphones to communicate and share information in a specific area.

Fig. 2: traditional visualization vs the proposed visualization (a) for people, (b) for restaurants and (c) for event directories

B. *Smartphones and user profiles*

The smartphones are used for two different tasks. Firstly, they exchange social information about their owners with the computational system. These devices can access a user's social networks that are today widely used and are convenient to provide relevant information about a user. The social data, available and collected through APIs or directly typed by the user, are stored in a smartphone application and used by this application to build the profile of the user. Fig. 4 shows that some social networks, like Pinterest, use such a technology to collect social data. A user's profile can be described as a collection of different social attributes that may also be divided into several sub-attributes. Fig. 5 describes how the profile of a specific user can be separated into different social attributes, for example, the identity, the preferences and the location history. The identity may be composed by a name, an age and a picture. The preferences can be divided again into several categories, such as sport and music. Such a profile is completed by the social information that is sorted into the different social attributes and sub-attributes. Secondly, the smartphones trigger the customized visualization. Since they can provide their precise location by using GPS or Wi-Fi technologies, their distance from the system can be known and a threshold can

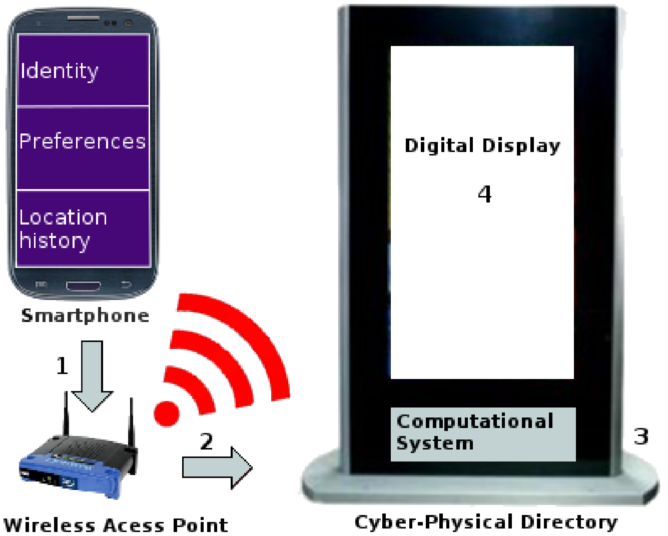


Fig. 3: System Architecture

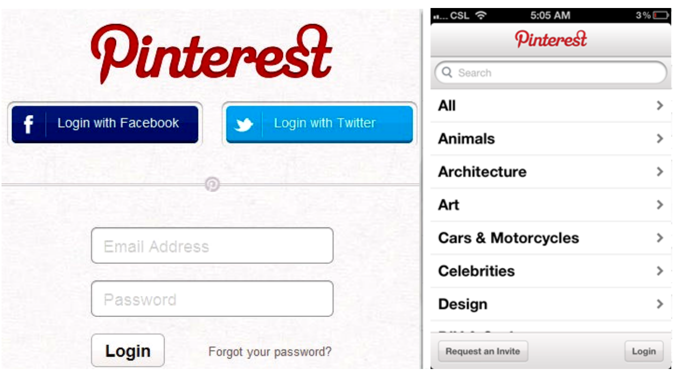


Fig. 4: Pinterest: two ways to collect social data a) by login via social networks b) manually

profile	social attribute 1	social attribute 2	social attribute 3
profile 1	identity 1 name, age, picture, ...	preferences 1 sport music ...	location history 1
profile 2	identity 2 name, age, picture, ...	preferences 2 sport music ...	location history 2

Fig. 5: A user as a combination of social attributes

be used to choose which smartphone is going to interact with the system.

C. The display device and the visualization

The display device is used to display a personalised and dynamic visualization from the computational system's results. It may be any kind of screen that can be linked to a computational system. In order to provide a customised representation for the

Fig. 6: Example of tagcloud

social things highly related to a specific user, a visualization based on tagcloud techniques is proposed. As presented in Fig. 6, a tagcloud gives greater prominence to words that appear more frequently in a source text and displays them by using different layouts [5]. Tagcloud is considered as an efficient visualization to find relevant information among large social data. For example, it can be used by a company to find interesting people for a job by displaying a bigger size for those qualified and experienced candidates [6]. Moreover, tagcloud may describe the social behaviour of a community by highlighting their common interests [7]. Here, the novelty is that the frequency of words used to create a tagcloud is replaced by the similarity between the profile of a user and the social things, such as events or places, previously stored in the computational system. In other words, originally the tagcloud only shows the keywords or certain social attributes larger as they are frequently accessed or mentioned; however, the proposed tagcloud visualization is used to show the social tie between a user and a piece of social information. Furthermore, the proposed visualization is dynamic. Some works have dealt with dynamic tagcloud visualization [8, 9], but in them dynamic means that the word cloud visualization is used to show an evolution over time. In this work, dynamic means that the visualization of the social things changes over time according to the users' closeness to the system. Finally, this visualization can also be adapted to different kinds of screens used to display information. On a digital screen, the desired information, i.e., information highly related to a user's interests, can be displayed bigger than other information. On a 3D screen, it could be displayed closer to the user. The main advantage of such visualization is that it highlights the relevant social information for a specific user and at a specific location. The tagclouds are created from www.wordle.net [10].

D. The computational system and the algorithm

The computational system is used for two different operations. Firstly, it stores the social things (events, places, people, etc.) displayed by the screen at a specific location. This information can be entered, modified or deleted by an administrator. As well, it collects the users' profiles sent by smartphones. Secondly, it runs an algorithm to find which social things are relevant to a specific user. As for the users, a profile can be created for the different things stored in the system, and this profile can also be a combination of different social attributes. Finally, users and things may both be described by a combination of social attributes, and the common social attributes can be used to find which things may be relevant to a user. Each social thing receives a score according to the similarity it shares with a user. The things with the best scores are shown with greater prominence to the user. It is the focus of the following algorithm.

1) The proposed similarity algorithm: The size and scale of a social thing are computed according to their relevance to a specific user. The algorithm is firstly presented in the case of a directory of people as described by Fig. 2(a), and then a generalization for every kind of social thing is proposed.

The Jaccard Index, also known as the Jaccard Similarity Coefficient, measures the similarity between two sample sets of data, A and B, as follows:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \tag{1}$$

Assuming that N people are displayed in the cyber-physical directory, a user u_l who is characterized by a profile l gets closer to the cyber-physical directory. This directory changes the visualization of the people in the directory according to his or her profile l. The new display is based on a similarity score provided by the computational system for each couple of profiles, (l, i), $i \in [1, N]$, $i \neq l$. Let us consider that F is the similarity function, the score S is calculated as:

$$S_l(i) = F(l,i) \tag{2}$$

where $i \neq l$. The higher the score $S_l(i)$ is, the more similar the profiles l and i are. In this example, dealing with a directory of people, F is calculated by the three social attributes described in Fig. 5: identity, preferences and location history. First, the Jaccard Index is computed for each social attribute and for each couple of profiles (l, i), $i \neq l$. Consequently, each couple of profiles is described by three similarity scores: $J_1(l, i)$, $J_2(l, i)$ and $J_3(l, i)$, which are functions of the profiles i and l. For example, the sets A and B in the Jaccard index of location history are the locations that two users visited respectively. The three similarity scores are normalized as follows:

$$j_n(l,i) = \frac{J_n(l,i) - \min(J_n(l,.))}{\max(J_n(l,.)) - \min(J_n(l,.))} \tag{3}$$

where $\min(J_n(l, .))$ and $\max(J_n(l, .))$ are respectively the minimum and the maximum of all the $J_n(l, .)$ for all $n \in [1, 3]$. This normalization places equal importance on each social attribute. F is a linear sum of $j_1(l, i)$, $j_2(l, i)$ and $j_3(l, i)$. In other words,

$$F(l,i) = W_1 j_1(l,i) + W_2 j_2(l,i) + W_3 j_3(l,i) \tag{4}$$

where W_1, W_2 and W_3 are the weights of each similarity score and $W_1 + W_2 + W_3 = 1$. The value of the weights are discussed in the next subsection. Finally, the size of the names or pictures of user u_i (for user u_l) is given by

$$Size_l(i) \propto S_l^2(i) \frac{S_l(i) - \min(S)}{\max(S) - \min(S)} \tag{5}$$

with $\min(S)$ and $\max(S)$ respectively the minimum and the maximum of all the $S_l(i)$ for $i \in [1, N]$.

The advantage of this similarity algorithm is that it may be generalized to many daily situations, as in those mentioned before. For a specific social thing, given N common social attributes (attributes in common in the user's profile and in the social thing's profile) and N weights, a score S_l can be computed by using:

$$F(l,i) = W_1 j_1(l,i) + W_2 j_2(l,i) + ... + W_N j_N(l,i) \tag{6}$$

where $\sum_{i=1}^{N} W_i = 1$ and $F(l, i) \in [0, 1]$. For example, for a cyber-physical directory displaying events, the different i can be the profiles of the different events and $j_1, ..., j_n$ the scores provided by the computation of the common social attributes. In this example, an event may be described by at least two social attributes, its location and its type. These attributes can be compared to the social attributes provided by the user: the location should be compared to the location history of the user, and the type to the user's preferences. Therefore, the cyber-physical directory can display a tagcloud of events according to the user's interests. More generally, this algorithm can be extended to many situations to provide an efficient visualization of information such that a user can find relevant information quickly.

2) The weights and the personalization: Two visualizations are considered: a personalized visualization and a default visualization. In the default visualization, the same weight is given to all the parameters. In the last case, that means that the values of W_1, W_2 and W_3 are the same. All of the social attributes have the same importance. In the personalized visualization, a user can choose the value of the weights on a specific scale between 0 and 1 to refine the parameters according to his or her current interest. For example, when this user goes to a new place, he or she may want to meet people who speak his or her own language, or if the user goes to an event, he or she may prefer to meet people who share the same interests. Therefore, the user may have the possibility to interact with the cyber-physical directory, by changing his or her profile manually, for example, or by changing the weights and refining the visualization. For instance, if the directory displays a tagcloud of events, the user may want to see only those that occur today. This personalization may be done through the smartphone application or by clicking on the screen, if possible.

3) Heterogeneous media data: The media data that interest a user can be in any form: people, events or others. Instead of getting only one kind of data at a time, a user may want to obtain heterogeneous media data from the system. The system can be extended to heterogenous media data easily. For example, when people and events are being processed and displayed, the system calculates S_l respectively. As S_l is ranged from 0 to 1, the size of the people and events listed can be based on S_l and Eq. (5). Two kinds of data could be shown on the screen based on S_l with a bigger size for contents that are more useful to the user.

III. EXPERIMENTAL RESULTS

The similarity algorithm is tested with two different approaches to collect users' profiles. The first approach is used to show how the system works with widely available data from social networks. The second approach proves the feasibility of a cyber-physical directory by implementing a web version of the system, where data are entered manually by the users.

A. User profiles through social networks

1) Dataset: The dataset [11] used in this paper contains information scraped from Foursquare. The advantage of Foursquare is that this social network can provide an identity, preferences of a user and his or her different locations over time. 213 users chosen randomly from the city of Los Angeles were extracted from this dataset to simulate the effect of physical location. From there, each user could be characterized by his or her nationality, friends, preferences and location history. Here, the preferences are the categories of places the user has visited, and the location history is the different locations that the user has been to. The nationality and friendships are grouped together because these two parameters are intuitively highly correlated. In the end, each user is described by three social attributes: identity, preferences and location history. The purpose of such a dataset is to show how the system and the similarity algorithm work with real social data in real situations. For a specific user, two visualizations are displayed and analysed.

2) Results: When the cyber-physical directory computes the similarity algorithm for a specific user, the visualization must be efficient to enable the user to see quickly who interests him or her. Moreover, the displayed people have to be similar to the user. To know the efficiency of the similarity function, the dataset is tested in both the personalized visualization and the default visualization. Fig. 7 presents, as an example, the results for a user called Etienne for the default visualization ($W_1 = 0.33$, $W_2 = 0.33$, $W_3 = 0.33$) and for a personalized visualization ($W_1 = 0.1$, $W_2 = 0.8$, $W_3 = 0.1$). In this case, the default visualization gives the same importance to identity, preferences and location history, and the personalized visualization gives greater interest to people sharing the same preferences as the user. Only the five most relevant names are displayed.

(a) not customized visualization

(b) with default weights

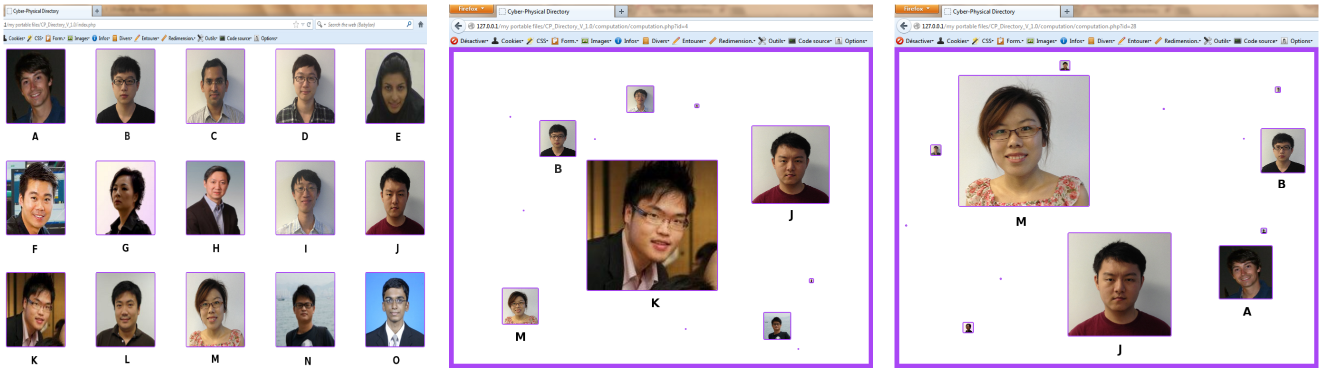
(c) with personalized weights

Fig. 7: 3 visualizations for a user, Etienne

According to Fig. 7, the layout of the cyber-physical directory shows that in the default visualization, "Julien" should be more attractive to "Etienne" than others. In contrast, "Juliette" is the biggest in the personalized visualization. "Juliette" and "Florence" are present in the top five for the two visualizations. "Juliette" shares eight preferences out of 21 with "Etienne". This score makes her the most interesting in the personalized visualization, but the fourth in the default visualization. This can be explained by the fact that, though "Juliette" and "Etienne" have preferences in common, their identities and location histories are quite different. "Florence" shares only six preferences out of 19 with Etienne, but her identity and her location history is more similar to that of "Etienne", so her name is displayed in the two visualizations. For the person called "Zachary", his name is only displayed in the personalized visualization. He shares 10 preferences out of 30 with "Etienne", but their identities and location histories are very different.

B. User profiles through manual inputs

1) Implementation: A web version of the cyber-physical directory has been implemented to prove the feasibility of the cyber-physical directory and of its algorithm. HTML, PHP, JavaScript and MySQL languages are used to build and run the system and the social information is collected from the website of the HKUST-NIE Social Media Lab. This social information about people working in the lab has been edited manually and sorted into three social attributes: identity, preferences and



(a) (b) (c)

Fig. 8: Implementation of the cyber-physical directory

location history. The same weight is given to all the social attributes. Fig. 8(a) shows a layout of the system before any manipulation. Instead of using the action of someone going closer to the system, a click on the pictures is used. This click triggers the algorithm to compute the scores for each user and displays a personalized visualization of people. The size of the pictures depends on the similarity between the user and each person.

2) Results: Fig. 8(b) and Fig. 8(c) are, respectively, the results of clicking "A" and "K". In Fig. 8(b) the picture of user K is the biggest one.That is coherent because K and A share a similar research interest and sport preferences. However, the image of A is not the biggest one in Fig. 8(c) when K is clicked. This is because K is more similar to M than A is. Moreover the picture of J is smaller in Fig. 8(b) than in Fig. 8(c). This is because, according to the profiles, even if A, K and J share similar preferences and location histories, K and J have the same nationality. As well, the picture of B is small in both Fig. 8(b) and Fig. 8(c). This is because B and K only have a similar identity, while B and A only share a similar location history. Finally, the system works and displays a personalized visualization that enables a user to quickly find relevant people.

IV. CONCLUSION

This paper presents a new cyber-physical system called the cyber-physical directory, which uses tagcloud techniques to enhance the visualization of social information according to a user's interests and preferences. By computing a similarity algorithm, it enables users to quickly find information that may interest them. The algorithm is tested with real data collected from Foursquare and with data directly typed by users. An implementation of the system is also built to show its feasibility.

Many extensions could be added to this work. First, it may be useful to create a learning algorithm which performs and improves the default visualization of the similarity algorithm. This would make the cyber-physical directory more accurate each time the user uses it. Secondly, it could be extended to a multi-user visualization system. When two or more people are in front of the cyber-physical directory at the same time, the display of the system could be different from that of an individual user. The information that may interest both users could be shown. As well, it could be interesting to consider the case where the system is composed of multiple screens.

V. ACKNOWLEDGMENTS

This work is supported by HKUST-NIE Social Media Lab., HKUST

REFERENCES

[1] http://www.itouchinc.com/default.html

[2] L. Sha, S. Gopalakrishnan, X. Liu and Q. Wang, "Cyber-physical systems: A new frontier," in Machine Learning in Cyber TrustAnonymous Springer, 2009, pp. 3-13.

[3] E. COMPUTING, "Cyber-physical systems," 2009.

[4] http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=503286

[5] C. Seifert, B. Kump, W. Kienreich, G. Granitzer and M. Granitzer, "On the beauty and usability of tag clouds," in Information Visualisation, 2008. IV'08. 12th International Conference, 2008, pp. 17-25.

[6] V. Gatteschi, F. Lamberti, A. Sanna and C. Demartini, "Using tag clouds to support the comparison of qualifications, rsums and job profiles," in Emerging eLearning Technologies and Applications (ICETA), 2011 9th International Conference on, 2011, pp. 57-61.

[7] M. A. Hearst and D. Rosner, "Tag clouds: Data analysis tool or social signaller?" in Hawaii International Conference on System Sciences, Proceedings of the 41st Annual, 2008, pp. 160-160.

[8] M. Dubinko, R. Kumar, J. Magnani, J. Novak, P. Raghavan and A. Tomkins, "Visualizing tags over time," ACM Transactions on the Web (TWEB), vol. 1, pp. 7, 2007.

[9] T. Russell, "Cloudalicious: Folksonomy over time," in Digital Libraries, 2006. JCDL'06. Proceedings of the 6th ACM/IEEE-CS Joint Conference on, 2006, pp. 364-364.

[10] J. Feinberg. Wordle. http://www.wordle.net/, Aug 2009

[11] J. Bao, Y. Zheng and M. F. Mokbel, "Location-based and preference-aware recommendation using sparse geo-social networking data," in Proceedings of the 20th International Conference on Advances in Geographic Information Systems, 2012, pp. 199-208.