Enterprise Wisdom Captured Socially

(Invited Paper)

Charalampos Chelmis
Department of Computer Science
University of Southern California
Los Angeles, CA
chelmis@usc.edu

Vikram Sorathia, Viktor K. Prasanna
Ming Hsieh Department of Electrical Engineering
University of Southern California
Los Angeles, CA
{*vsorathi, prasanna*}*@usc.edu*

***Abstract*—Data availability in online social networks as well as the business world has lately not been an issue. Vast amounts of data are being generated by social networking users in the form of informal interactions. What has been an issue, is the transformation of data into useful information, that in time and with appropriate processing becomes knowledge. In this paper we examine knowledge generation under informal social communications, based on semantically enriched user-generated data and associated metadata. We dynamically capture users' interests and expertise using such semantically enriched content. Knowledge networks of users emerge, exhibiting collective intelligence. To capture such collective knowledge, we propose a novel knowledge base paradigm, which seamlessly integrates information from multiple platforms and facilitates knowledge extraction, mining, discovery and inferencing. Using semantically enriched user profiles, we compute semantic similarity between users and content in a joint semantic space, driving numerous applications.**

I. Introduction

The Web is transforming. From a graph of static pages, it has rapidly grown into a medium in which users are creating, utilizing, distributing, and rating information. Online social networks and social media in particular have significantly revolutionized the way people are searching for information. Social browsing [1] has become a primary method by which users discover new content on the Web.

The business world has not stayed unaffected by this phenomenal transformation. Microblogging capabilities have penetrated the enterprise environment [2] providing a medium for users to share day-to-day operational knowledge and domain knowledge, discuss about problem solving, relevant emerging techniques, applications and technologies, trends, etc. Enterprise microblogging services mostly emphasize on the business perspective and therefore their content revolves around their main business and work culture, work practices, and everyday problems (technical or otherwise related to business).

Regardless of the latent incentives and random variables that drive social activity and populated content, online social networks and enterprise microblogging services share quite a few characteristics [3]. Structurally, both exhibit power-law, small-world and scale-free properties. Contextually, both demonstrate assortative mixing characteristics [4] with respect to lexical and topical alignment [5], [6].

Researchers have mostly focused on analyzing online social networks, even though social graphs have been mined out of numerous sources. E-mail traffic analysis [7], [8] has lead the extraction of unofficial social networks in enterprise context, in an effort to understand how information flow in the enterprise differs from online social networks. [9] argued that "*information extracted from e-mails could prove useful in a knowledge management perspective*", as it would facilitate expert and community identification.

Enterprises however do not solely rely on e-mail traffic to share information among coworkers. Microblogging services constantly gain ground, while more traditional media like SharePoint and Office Communicator are heavily utilized as part of question-answering and problem solving processes. Active Directory provides a formal structure for the bulk of e-mail and/or social traffic to follow. However informal communication through microblogging services span organizational charts. [10] presented an API for gathering and sharing interpersonal connections across multiple services and demonstrated its potential value with a comprehensive qualitative analysis.

Likewise, online social networks users do not solely rely on their social networks to share information with their friends. E-mails, chat services, and blogging, bookmarking and rating sites are some of the ways people share information with their social circles. Again, similarities between online social networks and enterprise microblogging capabilities become apparent.

Corporate microblogging services, on the other hand, act as facilitators of knowledge search and integration. Exchanged messages may contain knowledge in the form of solutions to particular problems, or may provide links to multiple external sources, like textbooks, research papers, FAQs, and best practice documents. Knowledge in such cases may not have been formally represented initially. Only when a specific question is being asked and an expert answer is provided, knowledge can be formally modeled and captured. However, informal communication is inherently noisy, both in terms of presentation (e.g. unstructured, ungrammatical text) as well as knowledge quantity and quality (e.g. personal status updates generally contain information but not knowledge). [11] investigated workplace relationships built between coworkers using microblogging services and determined interaction patterns that signal personal versus professional closeness between colleagues. We focus on both interaction types since knowledge can be generated in professional and personal contexts alike. Knowledge revolving around professional relationships is typically "different" than knowledge created on a personal context, but nonetheless being knowledge in both cases, which unless otherwise captured is lost.

In this paper we propose the integration of phenomenally heterogeneous information sources under a common framework so as to capture, analyze and utilize knowledge to drive numerous applications, such as expert identification and search, key influencer identification, targeted content delivery, information ranking and personalization, and recommendation systems. We propose an approach to automatically capture the ever changing users interests by monitoring their informal communication activities throughout the social web [12]: a) social networks status updates, and directed message exchanges, event co-attendance, and group membership, b) e-mail traffic for emergent, informal group formation and content mining, c) chat messages d) blog posts, e) bookmarks and f) ratings. Such sources of information can be seamlessly compiled together into **S**ocial Network as **K**nowledge **B**ase (**SKB**), forming dynamically changing user profiles, which can then be utilized to enable numerous capabilities such as semantic similarity calculation between people. SKB, when deployed in a corporate context, leverages collective knowledge, enabling collective intelligence capturing, preservation, management, and analysis. The main contributions of this paper are summarized as follows:

- We propose a generative process for knowledge, diffused under specific social context, based on information originating semantically annotated data from numerous sources.
- We propose a novel knowledge base paradigm built on top of social network, which enables integration of data, information and knowledge captured in the form of informal communications from numerous information sources, unstructured, semi-structured, and structured, such as e-mail traffic, blogging, conversational sessions etc.
- We define rigorous social metrics, which we use to compute semantic similarity of users based on annotated content and extracted knowledge, under the context of dynamic, informal, social interactions.

II. KNOWLEDGE BASE

A knowledge base is a system that enables domain knowledge collection, organization, and retrieval. The Artificial Intelligence community has widely used knowledge bases in order to represent knowledge using rich modeling languages. Based on the closed world assumption, knowledge bases have proven to be quite effective in capturing knowledge, inferencing new, previously unknown knowledge and providing advanced search capabilities [13], [14]. Open world knowledge bases however have proven to be computationally expensive as their inference mechanisms often become intractable [15].

Machine-readable knowledge bases [16] store knowledge in the form of logic rules, which describe the knowledge in a computer-readable fashion, thus enabling automated deductive reasoning. Semantic Web technologies have effectively advanced machine-readable knowledge bases by formally describing the structure of stored data (entity types and relationships between them) using formal schemata. [17] uses ontology based social network models to infer non-obvious relationships between nodes. Similarly, [18] proposes an architecture that enables semantic social network analysis, focusing however on the graph aspect of the problem only. Finally, [19] describes an architecture, which allows multiple, heterogeneous knowledge-based systems to cooperate in a partially structured social network. This work however is concerned with planning theory, as a mechanism to coordinate actions between agents.

Human-readable knowledge bases store knowledge in some type of human understandable format, permitting users to cooperatively capture, create, and organize, manage, and augment knowledge as a mean of information sharing [20], [21]. In an organization, they might store troubleshooting information, articles, white papers, user manuals, knowledge tags, or answers to frequently asked questions [22], [23]. Due to information storage as hypertext (with hyperlinks between them), classic information retrieval techniques are typically used to organize and search for information.

Irregardless of internal representation, knowledge has to be communicated to humans at the front end. An intuitive user interface should enable visual manipulation of social interactions and content analysis results jointly, enhancing overall user experience and enabling quick expert identification, as well as assisting in understanding the processes that drive knowledge and information flow between such experts through social interactions. On the other hand, to facilitate inferencing of new, previously unknown knowledge, stored information should be machine readable, enabling information processing with numerous techniques such as text analytics and probabilistic reasoning. We argue that user-generated content and metadata should be stored in a structured manner that will allow semantic information interpretation and facilitate machine learning over massive data analytics, in order to extract knowledge.

To benefit from well formed, structured knowledge, we propose storing data (similarly information, knowledge) in a semantic repository, which is linked to domain as well as external ontologies, folksonomies and vocabularies. OpenCalais[1], AlchemyAPI[2], and Evri[3] are examples of semantic information providers and annotation enablers, exposing rich APIs for text analysis and text annotation, entity identification, and topic discovery, as well as complex relationships mining. Linked Open Data[4] can further be exploited to gain insights into knowledge that may not be inherently present in the system under examination, but is accessible through external sources. Linking such external semantic repositories, elevates the knowledge base into an integrated, semantically rich heaven, where raw data becomes information, which after semantic processing drives knowledge inferencing and extraction.

Knowledge is generated, captured and shared without being limited to a specific language or system, but encoded in multiple formats, and distributed over various repositories. In large organizations, knowledge can be in the form of standard operating procedures, questioning and answering forums, FAQs, internal websites, social network, personal email communication and other means of communication. In such cases, knowledge is highly dynamic and constantly evolving, and unless otherwise captured it becomes "buried knowledge" [24]. To cope with management of such knowledge, conventional approaches of creating a single ontology may not be feasible or sufficient. Subsequently, the concept of traditional "knowledge base" also does not serve the purpose. Instead, the knowledge base has to continuously adapt to constantly progressing knowledge generation processes. We argue that "smart indexing" is required, allowing knowledge interpretation across information sources and facilitating complex query execution over an integrated view of the knowledge corpus. Instead of indexing knowledge using traditional information retrieval approaches (e.g. inverted indexes) or modern social techniques (e.g. tags or social bookmarks), dynamically and previously unmodeled knowledge can only be indexed and retrieved or sought in the context of social interactions based on informal communications, such as Q&A and resource recommendation.

Knowledge bases mainly focus on content, facts about it, and knowledge and reasoning on top of it, without considering the impact of social connectivity into knowledge discovery and sharing. On the other hand, social networking analysis is based upon the connectivity between users and resources, representing such relationships with social graphs. Graphs alone however, are able to neither represent different relationship types and levels nor to keep track of relationships length and communication frequencies, and are unable to capture contextual information associated with published content. Semantic Social Networking Analysis is actively trying to address the lack of semantics. However all approaches focus either on the underlying social connectivity graph or the content alone [25]. The multidimensional social network proposed in [26], which is closer to our work, uses an aggregated view of directed and inferred relations between users to perform user recommendation (i.e. link recommendation). Our focus instead is the creation of a social knowledge base which is capable of capturing and preserving knowledge generated under diverse contexts and considering informal interactions of multiple types, both direct and indirect. In that sense, link recommendation is only one of many features enabled by our framework. Moreover, in our work every dimension is significant depending on context and user need.

[9] provides a comprehensive summary of state of the art approaches for knowledge capturing from informal communication exchanges, and proposes an approach to facilitate expert discovery, which however focuses solely on e-mail exchanges. [27] proposes a framework for social annotation mining, which automatically identifies experts (i.e. knowledgeable users who create high quality annotations) and uses their knowledge to guide more detailed and accurate folksonomy learning in the photo sharing site Flickr. Interestingly, the authors conjecture that "*including annotations from non-expert, or novice, users leads to more comprehensive folksonomies than experts' knowledge alone*". In this work we consider the global sphere of users' activities by investigating interactions between phenomenally heterogeneous information sources, revolving around social networking activity, but not quite entangled to it. Informal communication (e-mail content and the pair of sender-receiver) undeniably provide hints on common interests and sharing of knowledge. On the other hand, social activities offer emergent semantics and allow expert identification to be conducted in a casual, informal context. We believe that richer and deeper understanding of such interactions can be further mined from numerous other sources and both worlds can be beneficial to one another when jointly considered together.

We argue that a paradigm shift is required to cope up with constant **knowledge generation**. From a **knowledge creation** point of view, a knowledge base should be cooperatively managed by a broad range of users who will constantly and dynamically contribute new knowledge, along with the context in which such knowledge is created. From a **knowledge representation** point of view, instead of focusing on comprehensive knowledge representation in specific (perhaps disjoint) domains, knowledge should be handled in the various forms in which it is created, yet be formally organized to permit easy retrieval and reusability. The challenge is to design a system that enables capturing, indexing, processing, updating, querying and retrieving knowledge as and when it is created and shared. From a **knowledge coverage** point-of-view such a system should be able to handle dynamic and highly-linked content as

[1]http://www.opencalais.com/
[2]http://www.alchemyapi.com/
[3]http://www.evri.com/
[4]http://linkeddata.org/

opposed to comprehensive self-contained references to static content. Finally, from a **knowledge access** point of view, we envision users browsing knowledge in a social manner, in a fashion similar to question/answering systems or topic based access systems that involve multiple users (experts or not). Instead of spending considerable amount of time searching for information, users should be able to quickly identify a small subset of expert users to guide them through their gradient descent into the knowledge space.

III. SKB: Social Network as Knowledge Base

Social networks can be seen as consisting of two independent yet strongly interconnected components: the network and the data that is being generated in it. Current social networking analysis techniques consider only a small fraction of the total information universe, focusing either on the network structure alone or users' profiles and interactions.

Social networks evolve when users "friend" each other. Friend-of links fail to capture the strength of association between users and explicit relationships between them. For example, two users may be computer programmers, but interested in PHP and Java respectively. In this scenario, linking users based on a specific programming language misses the latent relationship in the dimension of computer programming. To uncover such hidden information, one must examine the semantics of user activity and the network, along various dimensions. Typically, social networks capture relationships in a one-dimensional manner: two users are connected by a single edge carrying the generic "friend-of" label.

Information sources, exogenous to the social network, can significantly contribute to our understanding of users activities and interactions. We propose a formal modeling of SKB that abstracts the semantics of social network and exogenous information sources into an integrated, context-aware, multi-dimensional space, thus enabling the correlation of seemingly different domains so as to investigate them in conjunction, using the social network as the backbone infrastructure for knowledge preservation, indexing, and management, knowledge discovery and inferencing. SKB captures users dynamically changing interests, interactions and activities in a multi-dimensional manner, and provides a mechanism to measure the strength of association between users across dimensions.

A. Representation

We introduce a novel social graph representation, shown in Figure 1, which not only contains social links between users but also maintains integrated information regarding users dynamically changing interests and activities, as captured by their social network, e-mail traffic, chat messages, blog posts, bookmarks and ratings.

Social Layer is the glue of SKB. This is none other than the social network graph, which nodes represent users

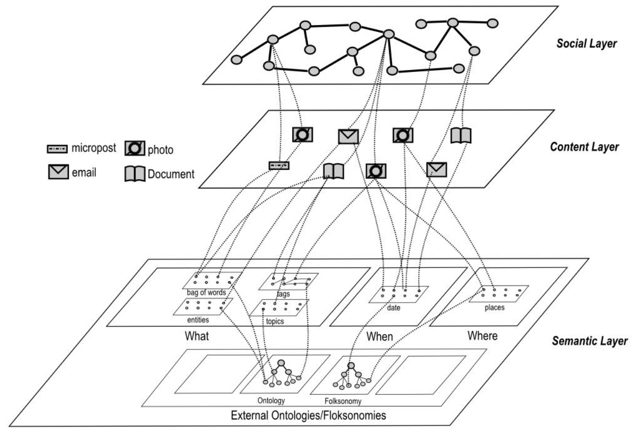

Figure 1. SKB Representation

and arcs represent explicit relationships (links) between them. To construct the social layer we start with friendship relationships from the social network and augment this initial graph by mining user relationships out of the rest of our information sources: e-mail correspondence, chat networks, blogging activity, shared bookmarks and common ratings. An edge between users is defined by the context under which it was created. Users may connect to others under multiple contexts (e.g sending e-mails and social status updates). The **Content Layer** captures published content from all available sources, including but not limited to resources shared by users (e.g. photos or videos), bookmarked and/or tagged resources (e.g. URLs), users' generated content (e.g. status updates in Facebook), e-mails, chat messages, and blog posts. The **Semantic Layer** contains meta-information about content, and can be broken into several constituting layers, each containing different metadata about content. This layer may include, but is not limited to, domain ontologies, folcksonomies, and taxonomies, external sources of formal knowledge, and linked open data. We call this stack of interlinked layers *enriched multi-layered social network*. This enriched social network enables analysis which spans layers, considering both multifaceted data and metadata, and the underlying informal communication graph. Knowledge is discovered, captured and inferred based on such complex information.

SKB users are completely being described by their content and associated metadata, given a context, each of which includes a variety of both textual and non-textual features, as shown in Figure 2. Some of the features are manually provided by users, while others are system generated. Social networks provide a variety of contextual features, that are dependent on the type of the resource which is annotated with. For example photos may have a geographic location attached to them while regular documents may not. However, many social networks share a core set of features. "*These*

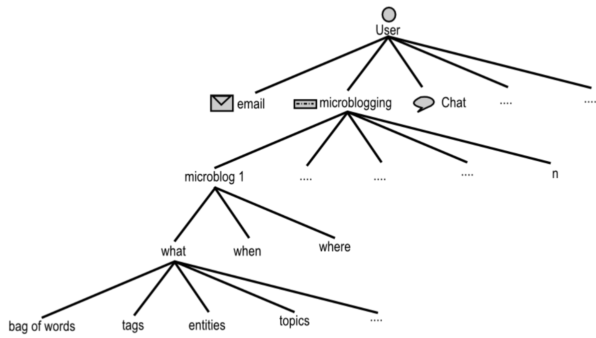

Figure 2. User representation in SKB.

features include: 1. author, with an identifier of the user who created the document; 2. title, with the "name" of the document; 3. description, with a short paragraph summarizing the document contents; 4. tags, with a set of keywords describing the document contents; 5. time/date, with the time and date when the document was published; 6. location, with the location associated with the document" [28].

We consider a representation of content by adapting the definition of event from [29], using each feature according to its type. Using this definition, content is completely described by a set of attributes (features), both textual and non-textual, that provide information regarding its "what", "when" and "where" aspects. Each user can in turn be completely described by her associated content. Next, we list the key types of features we use to model users in SKB.

- **Content Type**: This feature describes content type (e.g. document or e-mail) and its origin (e.g. social status update, e-mail body text or blog entry).
- **Textual Features**: These features include title and description (if available), raw textual content (bag-of-words), as well as user provided tags.
- **Enriched Annotation Features**: These features include resolved resources (i.e. URLs), named entities (elements belonging to a set of predefined categories, like persons and organizations), and topics (a general description of the topic(s) that content belongs).
- **Date Features**: Date values regarding content creation. We represent date values as the number of minutes elapsed since the Unix epoch.
- **Location Features**: Location metadata associated with content (e.g. geographical coordinates).

B. User Similarity

Based on user representation in SKB we are able to calculate similarity strength between any two object types in SKB, allowing us to compute semantic similarity, measure on a [0,1] range, between users and content in a joint semantic space. With respect to users, different similarity calculations can be performed to convey different meanings about how similar two users are, given a context/dimension. For instance to examine the amount of interests/knowledge shared between two users we can compute their similarity value with respect to their "what" dimension. We can further identify communities of users that exhibit collective knowledge by discovering linked groups of users with higher values of similarity among each other.

Next, we define rigorous social metrics, which we use to calculate similarity scores that span the layers of SKB. Even though we focus on user similarity, any two objects can be provided as inputs to our similarity metrics.

1) Social Dimensional Distance (ω_d): Given users x and y, dimension d and similarity measure $sim(x, y)$, we define **Social Dimensional Distance** as the similarity between users x and y along dimension d.

Definition 1 (Social Dimensional Distance):

$$\omega_d \doteq \delta_d(x, y) = \frac{1}{|x_d|} \sum_{i=1}^{|x_d|} sim(x_{d_k}, y_{d_k}). \quad (1)$$

$|x_d|$ denotes the cardinality of x and δ_d is a variant of Hausdorff[5] point set distance measure used to compare sets, from which we adapt for calculating users similarity. We normalize the similarity instead of using the min operator as used in the original Hausdorff distance metric, since we want to compute similarity between two users over all sub-dimensions. Like the original Hausdorff distance metric, Social Dimensional Distance is ***asymmetric*** with respect to users: $\delta_d(x, y) \neq \delta_d(y, x)$.

2) Social Contextual Distance (ω_c): We define **Social Contextual Distance** as the cumulative distance between two users under context c (a set of dimensions). The contribution of each dimension is regularized by a weighting factor α.

Definition 2 (Social Contextual Distance):

$$\omega_c \doteq \delta_c(x, y) = \frac{1}{|x_c|} \sum_{i=1}^{|x_c|} \alpha_i \omega_i, \quad c = \{d_1, d_2, \ldots, d_k\}. \quad (2)$$

3) Social Distance (Ω): We define **Social Distance** as the cumulative distance between two users across all their dimensions. The result is the normalized sum of two user's social dimensional distances. The contribution of each dimension is regularized by a weighting factor α.

Definition 3 (Social Distance):

$$\Omega \doteq \delta(x, y) = \frac{1}{|x|} \sum_{i=1}^{|x|} \alpha_i \omega_i = \frac{1}{|x|} \sum_{i=1}^{|x|} \alpha_i \delta_i(x, y). \quad (3)$$

4) Social Dimensional Neighborhood (θ_d): We define **Social Dimensional Neighborhood** of user x as a set of users whose distance from user x along dimension d is less than threshold γ.

[5] http://en.wikipedia.org/wiki/Hausdorff_distance

Definition 4 (Social Dimensional Neighborhood):

$$\theta_d \doteq \phi(x, d, \gamma) = \{u \mid \delta_d(x, u) \leq \gamma\}. \tag{4}$$

5) Social Contextual Neighborhood (θ_c): We define **Social Contextual Neighborhood** as a set of users whose distance from user x under context c (a set of dimensions) is less than threshold γ. The contribution of each dimension is regularized by a weighting factor α.

Definition 5 (Social Contextual Neighborhood):

$$\theta_c \doteq \phi(x, c, \gamma) = \{u \mid \frac{1}{|x_c|} \sum_{i=1}^{|x_c|} \alpha_i \theta_i \leq \gamma\}, \quad c = \{d_1, d_2, \ldots, d_k\}. \tag{5}$$

6) Social Neighborhood (Θ): We define **Social Neighborhood** as a set of users whose distance from user x across all their dimensions is less than threshold γ. The contribution of each dimension is regularized by a weighting factor α. Different thresholds may be provided for different dimensions.

Definition 6 (Social Neighborhood):

$$\Theta \doteq \phi(x, \gamma) = \{u \mid \frac{1}{|x|} \sum_{i=1}^{|x|} \alpha_i \theta_i \leq \gamma_i\}. \tag{6}$$

IV. Motivating Use Cases

We now discuss two applications that greatly benefit from SKB. In particular, we examine: a) expert identification (both individual and collective) and b) discovery of best practices (previously unknown/uncaptured knowledge). We note that the range of applications that SKB can successfully address is unlimited.

A. Context Aware Expert Networks Extraction from Informal Communication

SKB dynamically captures, analyzes and integrates users' interests and areas of expertise, tracking every day users' activities in multiple contexts, instead of relying on static user-generated profiles, which may be incomplete (sparse) and/or obsolete. By leveraging contextual information shared through informal communication and by utilizing multidimensional, semantically rich similarity metrics, our approach is able to identify well formed groups of "knowledge-based networks" [9] by computing semantic similarity between people spanning layers (dimensions). Instead of only computing content similarity between users, SKB further considers users' network proximity in order to highlight groups of users, which exhibit shared knowledge or collective intelligence.

SKB dynamically adapts to changes in users' interests and areas of expertise by constantly keeping track of user generated content, instead of only considering static user generated profiles, which may be incomplete (sparse) and/or obsolete. SKB does not only consider communication rates between users but also the semantics (context) behind such communication activities. Finally, instead of representing users as unidimensional vectors and computing user similarity using cosine or likewise similarity metrics, multidimensional similarity metrics, which accommodate richer and deeper similarity analysis that incorporate semantics and span layers, are provided.

The significance of expert identification can be better appreciated when applied to large and dynamic enterprise environments, where the number of people is relatively large. Newly employed professionals must quickly adapt to their working environment, hence they need to quickly determine key people they should contact in order to better understand their responsibilities, day to day operations and practices, etc. Similarly, people are constantly looking across team borders to identify current experts on specific topics, without knowing each other.

Consider the following scenario where user Alice is located in the offices of a multinational company at Palo Alto. Alice is trying to identify fellow co-workers to form a team about a project she is going to be working on the following next few years. Alice wants to discover people who have prior experience in Semantic Web Technologies, if possible specifically using Jena and OWL, and who have been consistently using Eclipse for development of Java based cloud computing applications. Alice tries both keyword based search and rule based retrieval, which both result in a (ranked) set of people who fulfill Alice's requirements but are scattered around the globe. Alice further indicates that she would like candidate co-workers to be located near her and have some reputation among the labor force. In this query, an individual's reputation is a parameter that is constantly updated based on social network interactions.

Complex queries like "Find experts in Semantic Web Technologies (Jena and OWL), who have consistently been using Eclipse for development of Java based cloud computing applications during the last 5 years, and who are residing in Palo Alto (next to my office)" are not inherently being supported by knowledge bases. Instead, engineers spend $40\% - 60\%$ of their time seeking information [30], [31]. Having a system that quickly indicates experts accross dimensions enables interdisciplinary cooperations that span organizational charts, lessens time spent in searching for experts and improves productivity. [32] underlines the importance of knowledge networks in the enterprise.

B. Discovery of Common/Best Practices

A common practice is a method or technique that has been consistently used by numerous people as a standard way of doing things and/or solving problems. Best practice is a method or technique that has consistently shown results superior to other approaches. Common practices can become best practices and best practices can become even better as improvements dynamically appear.

Documenting and charting procedures and practices is a complicated and time-consuming process, often skipped by companies, requiring significant amount of time to traverse the organizational chart for it to be reused [33]. The problem of knowledge capturing and preservation becomes even more important [34] with the retirement of senior experts from a corporation. Without proper knowledge capturing and preservation processes in place, cumulative knowledge is doomed to be lost.

Traditional knowledge management approaches overlook the knowledge utilization process, keeping no track of how knowledge is being effectively used. However, in a social networking context, it is possible not only to identify the utilization of knowledge by specific users, but also to associate to it additional context, in which the knowledge was employed. The expert who made the referral to a requested resource or hinted the solution can be identified, along with the question being answered and the problem being solved.

SKB facilitates discovery of common practices by examining the number of people who are actively using (or have used in the past) a method to solve specific problems that (may) span communities. For example, the choice between MySQL and Jena due to different requirements for different applications may be mentioned in different communities, both in a social context (social network), personal e-mail correspondence and chat exchanges. To quickly discover which database or semantic repository a user should use, she would probe the knowledge base for relevant solutions, along with the number of people who have used such solutions successfully in the past. To identify best practices, a user would specify the problem, along with a few positive keywords and the desired number of people to use as threshold in determining if a common method has received positive feedback.

V. Conclusion

In this paper we examined knowledge generation under informal social communications, based on semantically annotated user generated data. To capture and utilize dynamic knowledge that spans information repositories and data silos, we proposed a novel knowledge base paradigm, which seamlessly integrates information from multiple platforms and facilitates knowledge extraction, mining, discovery and inferencing from informal communications. Finally, we provided a methodology to compute semantic similarity of users based on extracted knowledge, under the context of informal social interactions. We examined two constructive applications which highlight the effectiveness and perceived benefits of SKB.

We argued that information extraction from informal communications can prove beneficial to enterprises that have adopted various forms of social networking services, like microblogging. We are currently building SKB, which we will use to conduct experiments on informal corporate networks and online social networks to quantitatively evaluate the effectiveness of our approach.

Acknowledgment

This work is supported by Chevron Corp. under the joint project, Center for Interactive Smart Oilfield Technologies (CiSoft), at the University of Southern California.

References

[1] K. Lerman and L. Jones, "Social browsing on flickr," 2006. [Online]. Available: http://www.citebase.org/abstract?id=oai:arXiv.org:cs/0612047

[2] J. Zhang, Y. Qu, J. Cody, and Y. Wu, "A case study of micro-blogging in the enterprise: use, value, and related issues," in *Proceedings of the 28th international conference on Human factors in computing systems*, ser. CHI '10. New York, NY, USA: ACM, 2010, pp. 123–132.

[3] C. Chelmis and V. K. Prasanna, "Micro-blogging in the enterprise: A few comments are in order," 2012, submitted.

[4] M. E. J. Newman, "Mixing patterns in networks," *Phys. Rev. E*, vol. 67, p. 026126, Feb 2003, http://link.aps.org/doi/10.1103/PhysRevE.67.026126.

[5] J. Weng, E.-P. Lim, J. Jiang, and Q. He, "Twitterrank: finding topic-sensitive influential twitterers," in *Proceedings of the third ACM international conference on Web search and data mining*, ser. WSDM '10. New York, NY, USA: ACM, 2010, pp. 261–270.

[6] M. McPherson, L. S. Lovin, and J. M. Cook, "Birds of a feather: Homophily in social networks," *Annual Review of Sociology*, vol. 27, no. 1, pp. 415–444, 2001.

[7] P. A. Gloor, R. Laubacher, S. B. C. Dynes, and Y. Zhao, "Visualization of communication patterns in collaborative innovation networks - analysis of some w3c working groups," in *Proceedings of the twelfth international conference on Information and knowledge management*, ser. CIKM '03. New York, NY, USA: ACM, 2003, pp. 56–60.

[8] J. Diesner, T. Frantz, and K. Carley, "Communication networks from the enron email corpus it's always about the people. enron is no different?" *Computational & Mathematical Organization Theory*, vol. 11, pp. 201–228, 2005.

[9] A. L. Gentile, V. Lanfranchi, S. Mazumdar, and F. Ciravegna, "Extracting semantic user networks from informal communication exchanges," in *Proceedings of the 10th international conference on The semantic web - Volume Part I*, ser. ISWC'11. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 209–224.

[10] I. Guy, M. Jacovi, E. Shahar, N. Meshulam, V. Soroka, and S. Farrell, "Harvesting with sonar: the value of aggregating social network information," in *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, ser. CHI '08. New York, NY, USA: ACM, 2008, pp. 1017–1026.

[11] A. Wu, J. M. DiMicco, and D. R. Millen, “Detecting professional versus personal closeness using an enterprise social network site,” in *Proceedings of the 28th international conference on Human factors in computing systems*, ser. CHI ’10. New York, NY, USA: ACM, 2010, pp. 1955–1964.

[12] A. Mikroyannidis, “Toward a social semantic web,” *Computer*, vol. 40, pp. 113–115, November 2007.

[13] T. Gruber, “The role of common ontology in achieving sharable, reusable knowledge bases,” in *Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference, Cambridge, MA*, 1991, pp. 601–602.

[14] N. Guarino and P. Giaretta, “Ontologies and knowledge bases: Towards a terminological clarification,” *Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing*, pp. 25–32, 1995.

[15] U. Hustadt, “Do we need the closed world assumption in knowledge representation,” *Working Notes of the KI*, vol. 94, pp. 24–26, 1994.

[16] R. Neches, R. Fikes, T. Finin, T. Gruber, R. Patil, W. Swartout *et al.*, “Enabling technology for knowledge sharing,” *AI magazine*, vol. 12, no. 3, p. 36, 1991.

[17] P. O. Wennerberg, “Ontology based knowledge discovery in social networks.” [Online]. Available: http://langtech.jrc.it/Documents/0509_Oezden_OntologyBasedKnowledgeDiscovery_JRC-FinalReport.pdf

[18] G. Erto, M. Buffa, F. Gandon, P. Grohan, M. Leitzelman, and P. Sander, “A state of the art on social network analysis and its applications on a semantic web,” in *Proc. SDoW2008 (Social Data on the Web), Workshop held with the 7th International Semantic Web Conference*, Karlsruhe, Germany, October 2008.

[19] D. A. Carlson and S. Ram, *Modeling organizations as a social network of distributed knowledge-based systems*. Los Alamitos, CA, USA: IEEE Computer Society Press, 1996, pp. 162–171.

[20] H. Hasan and C. C. Pfaff, “The wiki: an environment to revolutionise employees’ interaction with corporate knowledge,” in *Proceedings of the 18th Australia conference on Computer-Human Interaction: Design: Activities, Artefacts and Environments*, ser. OZCHI ’06. New York, NY, USA: ACM, 2006, pp. 377–380.

[21] A. Hester, “Innovating with organizational wikis: factors facilitating adoption and diffusion of an effective collaborative knowledge management system,” in *Proceedings of the 2008 ACM SIGMIS CPR conference on Computer personnel doctoral consortium and research*, ser. SIGMIS CPR ’08. New York, NY, USA: ACM, 2008, pp. 161–163.

[22] G. Leshed, E. M. Haber, T. Matthews, and T. Lau, “Coscripter: automating & sharing how-to knowledge in the enterprise,” in *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, ser. CHI ’08. New York, NY, USA: ACM, 2008, pp. 1719–1728.

[23] C. Wagner, “Breaking the knowledge acquisition bottleneck through conversational knowledge management,” *Information Resources Management Journal*, vol. 19, no. 1, pp. 70–83, 2006.

[24] V. H. Tuulos, J. Perkiö, and H. Tirri, “Multi-faceted information retrieval system for large scale email archives,” in *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, ser. SIGIR ’05. New York, NY, USA: ACM, 2005, pp. 683–683.

[25] C. Chelmis and V. K. Prasanna, “Social networking analysis: A state of the art and the effect of semantics,” in *Proceedings of the IEEE Third International Conference on Social Computing (SocialCom)*, October 2011.

[26] P. Kazienko, K. Musial, and T. Kajdanowicz, “Multidimensional social network in the social recommender system,” *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, vol. 41, no. 4, pp. 746 –759, july 2011.

[27] J.-H. Kang and K. Lerman, “Leveraging user diversity to harvest knowledge on the social web,” in *Proceedings of the IEEE Third International Conference on Social Computing (SocialCom)*, October 2011.

[28] H. Becker, M. Naaman, and L. Gravano, “Learning similarity metrics for event identification in social media,” in *Proceedings of the third ACM international conference on Web search and data mining*, ser. WSDM ’10. New York, NY, USA: ACM, 2010, pp. 291–300.

[29] B. Shevade, H. Sundaram, and L. Xie, “Modeling personal and social network context for event annotation in images,” in *Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries*, ser. JCDL ’07. New York, NY, USA: ACM, 2007, pp. 127–134.

[30] D. W. King, J. Casto, and H. Jones, *Communication by Engineers: A Literature Review of Engineers’ Information Needs, Seeking Processes, and Use*. Washington: Council on Library Resources, 1994.

[31] M. A. Robinson, “Erratum: Correction to robinson, m.a. (2010). an empirical analysis of engineers’ information behaviors,” *Journal of the American Society for Information Science and Technology*, vol. 61, pp. 1947–1947, September 2010.

[32] R. Cross, A. Parker, and S. P. Borgatti, “A bird’s-eye view: Using social network analysis to improve knowledge creation and sharing,” *Knowledge Directions*, vol. 2, no. 1, pp. 48–61, 2000.

[33] C. O’Dell and N. Ostro, *If Only We Knew What We Know: The Transfer of Internal Knowledge and Best Practice*. Simon & Schuster, 1998.

[34] D. W. De Long and T. Davenport, “Better practices for retaining organizational knowledge: Lessons from the leading edge,” *Employment Relations Today*, vol. 30, no. 3, pp. 51–63, 2003.