Enriching Employee Ontology for Enterprises with Knowledge Discovery from Social Networks

Hao Wu, Charalampos Chelmis, Vikram Sorathia*
Yinuo Zhang, Om Prasad Patri, Viktor K. Prasanna*
Department of Computer Science
*Ming Hsieh Dept. of Electrical Engineering
University of Southern California
Email: {hwu732, chelmis, vsorathi, yinuozha, patri, prasanna}@usc.edu

***Abstract*—To enhance human resource management and personalized information acquisition, employee ontology is used to model business concepts and relations between them for enterprises. In this paper, we propose an employee ontology that integrates user static properties from formal structures with dynamic interests and expertise extracted from informal communication signals. We mine user's interests at both personal and professional level from informal interactions on communication platforms at the workplace. We show how complex semantic queries enable granular analysis. At the microscopic level, enterprises can utilize the results to better understand how their employees work together to complete tasks or produce innovative ideas, identify experts and influential individuals. At the macroscopic level, conclusions can be drawn, among others, about collective behavior and expertise in varying granularities (i.e. single employee to the company as a whole).**

I. Introduction

As social media have become phenomenally popular, enterprises have adopted light-weight tools such as online forums and microblogging services for internal communication [32]. As such, employees have been using social network sites and microblogging services in the enterprise for various reasons: to stay in touch with close colleagues or to reach out to employees they do not know, to connect on a personal level or establish strong professional relationships in order to advance their career within the company [6]. Others, perceive the use of such services as extra communication channel for news reading and company events notification, a mean to promote their ideas, contribute to conversations revolving around company matters, or participate in discussions of work issues. In this sense, user activity and behavioral data in this context contain valuable information. User's interests in a personal or professional level can be discovered, whereas interesting communication motifs can be mined out, enhancing our understanding on employees' communication patterns as well as patterns of information propagation and browsing in enterprise social networks.

Enterprises can utilize the results stemming out of enterprise social media services, among others, to better understand how their employees work together to complete tasks or produce innovative ideas, identify experts and influential individuals so as to evaluate and adjust their management strategy, team building and resource allocation policies. Similarly, employers can benefit from enterprise social media services analysis in multiple ways. Recommendation services can provide better results in terms of "interesting" people to connect to, as well as suggest "interesting" discussions for employees to contribute or projects to get involved in. Information filtering algorithms can better promote a subset of news instead of directly delivering all sorts of irrelevant data to employees, alleviating information overload from them, and enabling them to focus on information that does matter. Information acquisition, such as search for people, data and answers to problems can be significantly sped up, resulting in increased productivity through collaboration and problem deduplication. The usage of social media in enterprises in turn can be improved.

Social media analysis is a broad topic. In this work, we primarily focus on capturing employee's interests and areas of expertise as well as mining interconnections between employee's work-related activities and their social interactions on collaboration platforms used in working environments. Identifying employee's interests and areas of expertise is not a trivial task. Existing methods model user interests based on static properties or by keeping track of their collaborative activities. However, profiles of users may be completely unavailable or extremely scarce since users do not often populate enough information to describe themselves, and may be obsolete if users do not constantly update them to match their most up to date interests. On the other hand, high-volume activity on enterprise social media offers a unique opportunity for content analysis, as active participation of employees and sharing in informal conversations makes it possible to identify knowledge and expertise by analyzing messages that users post. Nonetheless, content from collaboration activities may be significantly short (e.g., 140 characters in Twitter) and inherently noisy. Often microblogging content does not adhere to any grammatical or syntactical rules, contains slang terms, user-defined hashtags and emotions or other special characters, which denote emotions or user-defined notions, the semantics of which can be unknown or not previously modeled.

In this paper we present an approach to enhance collaboration analytics [30]. We propose an employee ontology that captures user behavior within a given context as a result of robust statistical learning analysis. We leverage our ontology with a semantic-rule based methodology to provide insights with respect to corporate collaborative activities, and support complex semantic queries, enabling numerous applications, such as recommendation services, information filtering, semantic, and social matchmaking. Our solution consists of three steps: First, we build an ontology to capture static or rarely changing information about individual employees with respect to the formal organizational structure, past projects and working experiences. Secondly, we take advantage of the dynamic environment of communication platforms enterprises adopt, to mine employee's knowledge in multiple contexts. Even though we only discuss internal social media in this work, our approach is extensible to multiple other communication channels (e.g., e-mail), facilitating integration across formal and informal interactions of any sort. We extract employee's specialization skills and levels of expertise from users' activity in such communication channels. Finally, we populate our ontology with instances covering all aspects of employee's electronic activities described above. We use our semantic repository to facilitate complex queries involving concepts and relations between multiple contexts.

Our contributions can be summarized as:

- We enrich the existing employee ontology to cover aspects such as expertise, skills, and prior experience, which are useful in business processes.
- Our ontology captures and preserves knowledge, which is mined from informal communication at the workplace. We apply statistical learning methods to leverage collaborative content, uncovering trending discussed latent topics and employees' expertise.
- We demonstrate how our ontology can be utilized to provide analysis and insights both at a micro and a macro level with complex semantic queries that result among others in expert identification (micro level) and corporate-wide collective wisdom (macro level).
- We present a case study based on a real-world, large-scale dataset from a Fortune 500 company.

II. Related Work

We focus on enriching employee ontology in enterprise context, where employees are the users of social media in workplace. Hence, our study is essentially related to user profile ontology [12], which is application-specific and can support multi-dimensional user-centric features. It has been proposed to apply to various domains such as web search [28], peer-to-peer mobile system [29], social tagging services [16], personal information management [11]. The research of user modeling can trace back to decades ago. Rich [21] and Kobsa [13] presented overviews of methodology and related guidelines to create a user profile. We build on prior work, including in our ontology static employees' properties such as ID and position in the company.

Recent work on building context-aware ontologies [3], [31], [23] has focused on supporting changes of usage, activity, and resource. Context in this case may refer to persistence of properties, i.e. permanent or temporary, or evolution, i.e. static (such as user identity and universal preferences) and dynamic (such as activity, motion, state and orientation) [20], [25]. With the objective of reducing the human intervention for situation changes, Stan et al. [23] proposed a user profile ontology for situation-aware mobile services. The ontology contains concepts of real-time detections of situation changes and activation of user preferences for each situation. Interests [28], [10], [27] and expertise [9] are two of the most important factors considered in user profiling for most applications such as personalization of information acquisition. We use similar concepts in dynamic contexts in our ontology and fill in the knowledge needed to enable complex semantic search. Contrary to previous work, we discover knowledge by leveraging message conversations in enterprise social media, and characterizing areas of expertise and interests for individual employees using latent topic descriptors. Our ontology integrates multiple communication media, both formal (i.e. reporting relationships) and informal (such as e-mail and online social interactions), facilitating a broad range of complex semantic queries as a result. We discuss typical use cases in Section VI.

Knowledge discovery and semantic analysis in social networks has attracted much research attention in recent years [8]. A typical theme in this area is to discover the latent topics of the large-scale user-generated text corpus, where topic models [24] show their merits. They are so-called generative models which simulate the generation of words in documents [4]. Recent studies developed approaches that can be applied to challenging mining tasks including understanding relational links between documents [5], influence analysis of scientific articles [19] and in web blogs [18]. Another theme closely related to our study is expertise discovery in popular social media [14]. Various types of social networks have been investigated for automatically user expertise identification, including web forums [33], knowledge-sharing websites [2] and academic social networks [17], emails and chat logs [7]. In our study, we identify employees' expertise and interests by analyzing the message content of enterprise social media, specifically light-weight enterprise microblogging services which contain dynamic and timely information. Our ontology is extensible, permitting integration of multiple heterogeneous information sources, allowing their simultaneous analysis and mining.

III. Dataset

We investigate a large-scale dataset, which is a complete snapshot of an enterprise microblogging service used by employees in a Fortune 500 multinational company. The functionality of the microblogging service resembles that of Twitter, whereas its interface is similar to Facebook. The enterprise microblogging site does not impose any restrictions on the way people interact or who they chose to follow, much similar to Twitter, and its main purpose is to promote and enable collaboration and sharing within the company.

Similar to Twitter, users author messages in the enterprise microblogging service, and form threaded discussions. A message may be available to the corporate-wide news stream, sent to a specific group of employees, or be a direct message to a single individual. Each message may have been annotated with hashtags and may receive multiple replies by other employees. A threaded discussion may therefore be followed by its chain of replies. Message interactions in prompt microblogging may help resolve technical problems and notify team activities in work practices in a timely manner. Figure 1 depicts a typical scenario of enterprise social interactions. Employee_AABF participates in a project along with his colleagues. In performing his role, he comes across a problem and posts a question at the enterprise microblogging platform, where Employee_AAAD and Employee_AADE read the question. Employee_AAAD is interested in the topic of

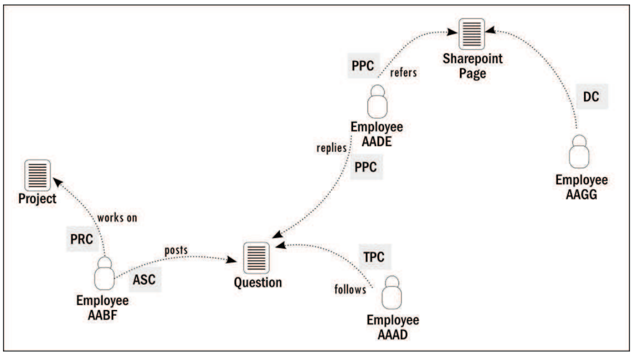

Fig. 1: Toy example of contextual enterprise social interactions

the question and starts following the question, while Employee_AADE responds to the question using a reference to a Sharepoint page, which was contributed by Employee_AAGG. Such interactions reveal collaborative, contextual relationships between employees, whereas the content of discussions reveals expertise and interests of individual employees.

The dataset we use includes $9,855$ unique employee users, which represent about 15% of the employee population. We collected all $15,200$ messages with explicit reply links to other messages over a time span of 1.5 years. We have cleaned the content of messages applying a stemming step, after which we obtain $4,384$ distinct words across all messages. Apart from collaborative activities, our dataset further contains organizational information about the employees who have joined the corporate microblogging service (i.e. a subset of the complete company hierarchy). For each employee we build a profile that maintains information regarding her current (and previous) position in the company hierarchy, as well as a historical profile of her past experience, project participation etc. For confidentiality and privacy protection, we anonymize the identifying information of users, organization levels and messages content. Each employee user is represented using a 4-character ID (e.g., "AECF"). The organization level of each user is denoted using a character from "A" to "M".

IV. Knowledge Discovery

We mine knowledge from message interactions between employees in the enterprise microblogging service. Recent advances in machine learning have developed statistical learning methods called topic models [4], [24] to automatically uncover latent "topics" in natural language texts. Topic models take advantage of co-occurrence of words in different text documents. They use hierarchical Bayesian models to capture the generation process of words in documents by introducing an intermediate latent topic layer. Each topic is represented as a mixture of words with probability distribution and a document can be decomposed into a distribution of various latent topics.

In our study, we adopt the Author-Recipient Topic model (ART [15]) to discover topics in messages posted by employees in the internal enterprise microblogging service. The model builds on Latent Dirichlet Allocation [4] and the Author-Topic model [22], but it's more powerful to capture the communication between authors. ART makes the assumption that each word w in a message is sampled from a multinomial distribution ϕ_t (the word mixture for a topic t). The topic is drawn from a multinomial distribution θ_{ij}, which is the topic mixture specific to the author-recipient pair (i, j) of the message. Instead of only modeling topic distributions over messages or authors, the ART model conditions the distribution over topics on both the author and the recipient of a message. Therefore, topics are discovered in a more natural way by modeling the interaction relationships between people. Furthermore, ART can capture not only how often employee pairs interact, but also which topics are more prevalent in their discussions. Applying ART in multiple communication channels further enables analysis of prevalent topics across mediums. One can then check if employees are consistently exhibiting expertise regardless of medium or if instead, they choose particular platforms to contribute and share, such as email correspondence, while remaining inactive in others.

A. Model Fitting

We train ART using Gibbs sampling [15]. We fix the hyperparameters $\alpha = 50/T, \beta = 200/V$, and set the number of topics $T = 100$. The Gibbs sampling procedure iteratively draws recipient j and topic t for each word w from the updated conditional distribution of it given all other words' topic and recipient assignments

$$P(i,t|\cdot) \propto \frac{\alpha + n_{ijt} - 1}{\sum_{t'}(\alpha + n_{ijt'})} \frac{\beta + n_{tw} - 1}{\sum_{w'}(\beta + n_{tw'}) - 1}, \quad (1)$$

and the parameters θ and ϕ can be estimated by

$$\hat{\theta}_{ijt} = \frac{\alpha + n_{ijt}}{\sum_{t'}(\alpha + n_{ijt'})}, \hat{\phi}_{tw} = \frac{\beta + n_{tw}}{\sum_{w'}(\beta + n_{tw'})} \quad (2)$$

B. Expertise and Interest

We use the trained model to capture not only latent topics, but also employees' "Expertise" and "Interests" from their microblogging activity. We assume the more an employee has replied to others to offer answers on a particular topic, the more knowledgeable the employee is. We note that other prominent indicators of expertise and external factors contributing to expertise should be considered, which however we leave as future work. To measure employee i's expertise on topic t, we aggregate the number of words n_{ijt} assigned to topic t and author-recipient pair (i, j), resulting in $\sum_j n_{ijt}$. Hence, we are able to quantify the expertise level each user i has on a specific topic t:

$$EL(i,t) = \frac{\sum_j(\beta + n_{ijt})}{\sum_{i'}\sum_j(\beta + n_{i'jt})} \quad (3)$$

We quantify the result to a discrete scale of 1 to 5. The larger the value, the more knowledgeable an employee on a topic. Expertise is therefore a relative measure that represents how proficient one is on a topic t, as compared to other employees. $EL(i, t)$ for different topics t for each user i reflect i's expertise sharing activity, and could be used as a complement to expertise described in resume and employee profiles. By conditioning on a particular topic t, we are able to rank "experts" based on their $EL(i, t)$ values. Such experts can be recommended as the recipients (e.g., to mention "@") when one would like to ask a question on a specific topic.

To measure employee i's interest on topic t, we use the α smoothed normalization and obtain

$$Interest(i,t) = \frac{\sum_j (n_{ijt} + \alpha)}{\sum_t \sum_j (n_{ijt} + \alpha)} \tag{4}$$

This is a relative measure as well, and represents to what degree an employee prefers one topic t to other topics.

At micro level, the knowledge discovered can be used to better understand the expertise areas and interests of individual employees and their collaboration behavior, so as to adjust management strategy, team building and resource allocation policies. Employees can benefit from the perspective of information personalization. Moreover, we emphasize on the dynamism of the knowledge discovered. Employee interests, skills and expertise can change depending on time, work orientation and responsibilities, project focus and overall team competence. Given a context (e.g., a group discussion versus a status update) may yield significant, different aspects of employees' focus. Depending on personal or professional nature of content different interests can be mined and different expertise levels can be identified for disjoint set of topics. Moreover, employees often assume multiple roles in multiple projects (e.g., an employee might act as manager in one project, while being a software developer in another). As a consequence employee interests, skills and expertise are both context and time dependent: can differ at multiple points in time, and be different at the same point in time yet within the boundaries of correspondence with different individuals.

At the macro level, companies can issue queries with respect to collective intelligence, interests and expertise, per department or across companies, aggregating results from individual employees. Trending topics can be mined out in varying granularities ranging from the atom (i.e. single employee) to the company, as an organization build by atoms, and everything in between.

V. Employee Ontology with Social Knowledge

In this section, we present a detailed description of our employee ontology, which we name Employee Ontology with Social Knowledge (EOSK)[1]. EOSK captures the ever changing, "contextual" notion of user interests, skills and expertise. The main reason we adopt ontology to represent employees and related business concepts is that it provides a generic, reusable, and machine understandable model required for describing business activities and measuring employees' behavior.

A. An Overview of Classes

From generic classes to derived classes, we create employee ontology in top-down manner to give a deeper description of the relations between concepts. Professional `Position`, `Expertise`, participated `Project`, and `Connections` for different `Contexts` are the foundations of our employee ontology. As we discussed above, one important feature we incorporate into our ontology is extracted knowledge. We extracted `Topics` and implicit `Expertise` from an internal microblogging service.

[1]We construct and edit the ontology using TopBraid Composer http://www.topquadrant.com/products/TB_Composer.html.

B. Properties and Interrelations of Classes

We now elaborate on the hierarchical relations of the classes and their properties. Figure 2 shows the panorama of classes in our ontology, which defines a structured set of concepts to characterize employees in comprehensive dimensions.

`Business_Associate` is a basic unit may have `Connection` and it is inherited by `Employee` and `Organization`. We consider both of them as entities that have connections between one another. `Employee` can have various types of connections, which come from different contexts. For example, each employee can have her/his direct supervisor as hierarchical connections; Employees may have email connections, and those who they have interacted with in internal online media. Organizations (e.g., a corporation) have employees as their members, and can be connected to other organizations or employees. Allowing organizations and employees to collaborate, we blur the difference between them and denote them as one fundamental unit, i.e., node in graph, thus building complex social networks. An `Organization` has its distinct identity "OrganiztionID", and it is an integration of `Companies` with various `Departments` and `Facilities`.

`Employee` is the core of our ontology. It is essential to include business concepts that are important for human resource management, such as identity "EmployeeID" and `Connections`. Note that the `Connections` object property of `Employee` are generated from `Business_Associate`. Different from user profiling [12], [26] where the ontologies mostly focus on static properties such as personal information, we emphasize on the usage of our employee ontology in dynamic enterprise contexts. Employees have `Position` in the enterprise and each participate in one or multiple `Projects`. The arrangement of each employee's `Position` should be aligned with the `Expertises` she/he has. The `Projects` that an employee participates in record the working status of the employee. We use "StartDate" and "EndDate" to describe the life span of the `Project`, based on which we can trace the state of projects. Moreover, we can assign a new project to an employee without conflicting the timeline of another project.

The `Position` class describes the current status of an `Employee` in the enterprise, containing information such as in which `Department`, `Facility` and `Company` the employee works. The exact employee position title (e.g., manager, senior engineer or CEO) is linked by "hasType" to the object class `Employee_Position`. To represent the hierarchical level of the employee in the company, we use "Orgnization_Level". The temporal aspect has been taken into account in `Position`. We use "EffectiveDate" to denote the starting date of an employee at this position and "ExpiryDate" to denote the date the current position will end. "ExpiryDate" can be updated, for example, when the employee continues the contract on the position. We can also design a mechanism to send notification messages when an "ExpiryDate" approaches. Based on the timelines of the `Positions` the employee work at, we can trace the working history of the employee for references in order to assign an new position. To assign new projects, semantic rules that ensure that time span of new projects won't exceed the current position "ExpiryDate" may be appropriate.

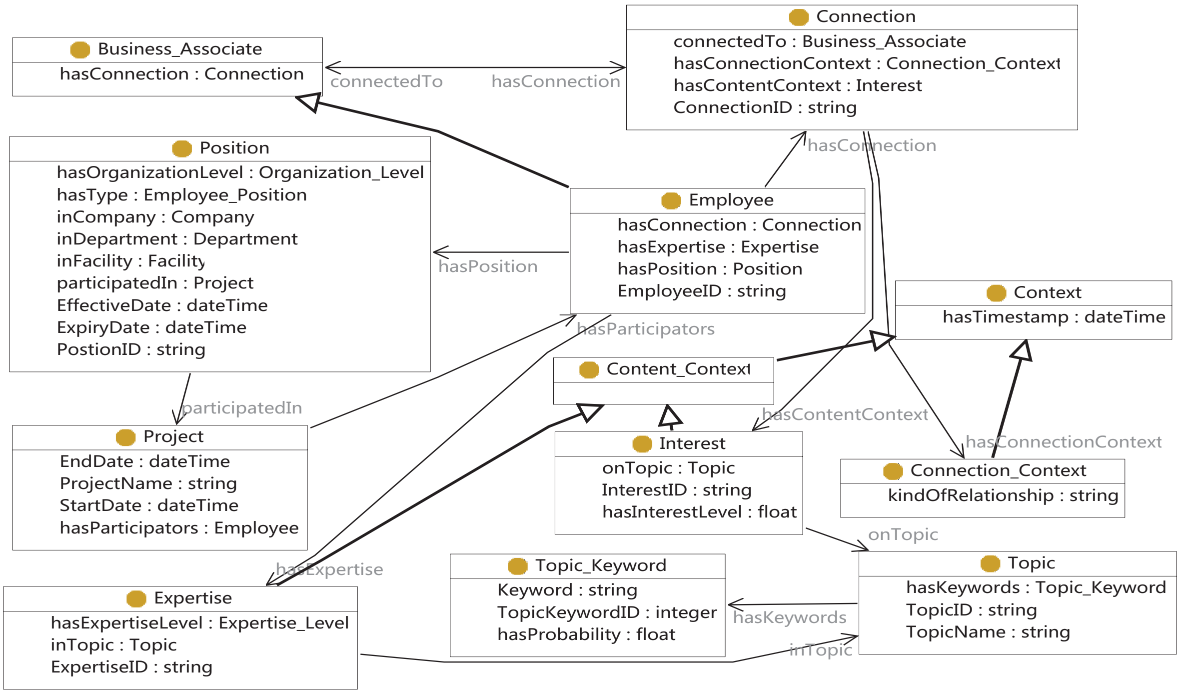

Fig. 2: A panorama of EOSK classes, properties and interrelations.

Expertise of an employee can be summarized from employee resumes and profiles from past appointments, if it does not violate privacy policy. It is difficult to measure expertise of each employees objectively in terms of different domains, skills [9]. An more challenging task is to identify expertise levels of each person with regards to each expertise area. In `Expertise` class, we define "ExpertiseID" and "Expertise_Level" to represent those concepts. We leverage social media the enterprises use as a knowledge base to infer the expertise. Peer validation in the form of informal interaction acts as supporting evidence of the level of expertise an individual possesses. We use the concept `Topic` (e.g., computer technology, management, accounting) in general to represent the expertise area. We take advantage of the message texts in enterprise microblogging and use a robust statistical learning method [15] to mine the individual expertise, as we discuss in section IV.

`Connections` refers to other people an `Business_Associate` (`Employee` or `Organization`) have in different social networks. Each business associate is connected to other ones in different contexts, which we define as `Connection_Context`. The `Connection_Context` specify what kind of relationship each pair of entities has. It can be colleagues, collaborators, or dominance/subordinate, etc. Each connection also has `Content_Context`, which represent the common `Interest` the two entities have. For example, in enterprise microblogging system, an employee user can be connected to other users with follower/followee relationship or simply reply relationship regarding to message exchange. They may have common preferences in discussing `Topics` such as computer technology. Next subsection, we discuss the generalization of `Expertise` and `Interest` as `Context`.

`Context` can be defined as any information that can be used to characterize the situation of an entity [1]. Context can be physical environment such as location, time and surrounding facilities, or the computing environment such as network , accessible devices to the entity [23]. In our employee ontology, we emphasize on the social environment that one employee may interact with others in various working situations. To better adapt the ontology to context-aware situations, we take into consideration of temporal feature and add timestamp to each context. `Context` is inherited by two classes `Content_Context` and `Connection_Context`. We may consider basic `Connection_Context`, such as email and social network types, which are the most important ones in workplace. We define `Expertise` and `Interest` as specific types of `Content_Context`. This captures the idea that both of the two class are related to `Topic` and have their own levels or strengths.

`Topic` refers to semantic concept that conveys specific ideas in a theme or subject [4], [15]. It could be the main ideas of a paper, article, newspaper, or human conversation. We use `Topic` to represent the areas or themes of each `Expertise` and `Interest`. We use the word can mostly convey the concepts to represent the "TopicName" of a topic. Each `Topic` has a list of `Keywords` with probabilities to represent to what extent the keyword conveys the semantics of a topic. Topics are uncovered using ART as we mentioned in Section IV.

VI. Case Study

In this section, we present a case study of complex semantic search based on our large-scale enterprise microblogging dataset[2]. We demonstrate the effectiveness of our proposed

[2]Instances are imported into Vituoso server http://www.virtuoso.com/

ontology through a number of use cases that exhibit multi-dimensional semantic search.

We design queries that can be used in different enterprise contexts. Such queries can be driven by individual employees to find for example information of interest, or managers or human resources personnel to identify experts and effectively assembly teams for specific tasks. To personalize individual needs in message reading and interaction, it is necessary to identify individual interests and content context he/she likes to participated in. From an administrative point of view, it is also important for enterprises to manage the human resource and assign each employee on positions that match individual expertise and interests.

Here is a motivating example. Suppose you are a manager in charge of a newly arrived project. You need to find someone who has the expertise to work on the project. Moreover, when planning and decomposing the tasks of the project, you have many specific technical problems to ask someone else to solve. It also needs semantic search of people who might implicitly have some kind of expertise on a specific topic. Referring to each employee's resume might not be a good way since there is usually high-level description of expertise, and resume might be unavailable due to privacy issue. Moreover, to build the team, some employees who have had close collaboration might be a better fit for the coherence of the team. This requires complex semantic search on expertise and relationships between employees. With discovered knowledge, EOSK captures the query semantics in these cases. In the following, we present typical queries using SPARQL and the results from our intuitive user interface, that is developed using ASP.NET.

A. Expertise Identification

To fill the positions of a new project, one may have to find someone with relevant expertise. The quantification of expertise levels in different areas for each employee is critical for position assignment and team initialization. As we discussed in Section IV, we mine employees' expertise areas from their informal interactions. Their levels of expertise may vary from topic to topic. For instance, an expert in computer technology might know little about lawsuits. A micro level query that retrieves topics and levels of expertise for a given employee follows.

Query 1: Given an employee, get his expertise topics and expertise levels

```
PREFIX EOSK:<http://../Employee.rdf#>
SELECT ?Expertise ?Topic ?Level
WHERE {
  EOSK:Employee_AABW EOSK:hasExpertise
?Expertise.
  ?Expertise EOSK:hasinTopic ?Topic.
  ?Topic EOSK:hasTopicName ?Name.
  ?Expertise EOSK:hasExpertiseLevel
?Level.
}
```

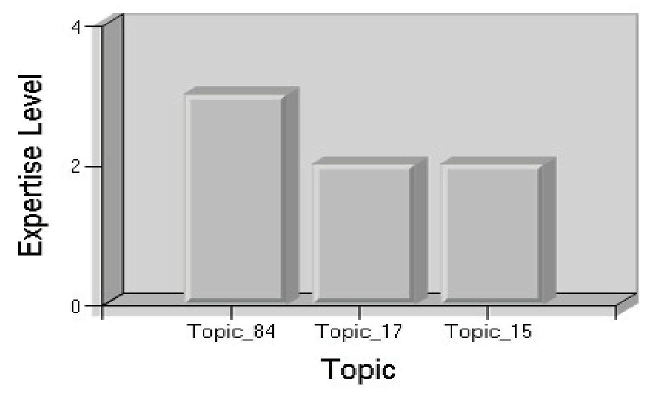

Fig. 3: Sample SPARQL query for expertise identification and search result.

B. Collaboration Across Hierarchy

Organizational hierarchy is static, and dated, whereas communication may reflect "shortcuts", i.e. collaboration that spans hierarchical levels. Seeking help, offering guidance, acting as mentor, thought leaders, influencer etc. Study of communication may reveal change in organizational dynamics. For instance, employees Employee_AABG and Employee_AACD may appear as team-members according to organizational hierarchy, but reality (i.e. real life informal interactions between them in the corporate microblogging service) may hint otherwise. Our intention is to better understand how information propagation works between corporate borders, and identify potential shortcuts in the organizational chart. In the case of new project arrival, one might need to find employees at different hierarchical levels (e.g., low-level software engineer, senior technician, manage, even CTO) to fill the position and work together in a team. To this end, we may examine employees' connections under organization hierarchy connection context and compare these to "connections" due to interactions in the microblogging service. The query is shown as follows.

Query 2: Given an employee, retrieve all his connections and their level in organization hierarchy

```
PREFIX EOSK:<http://../Empoyee.rdf#>
SELECT ?Connection ?Employee ?Position
?Organization_level
WHERE {
  EOSK:Employee_AABW
EOSK:hasConnection ?Connection.
  ?Connection EOSK:hasconnectedTo
?Employee.
  ?Employee EOSK:hasPosition
?Position.
  ?Position EOSK:hasOrganizationLevel
?Organization_Level.
}
```

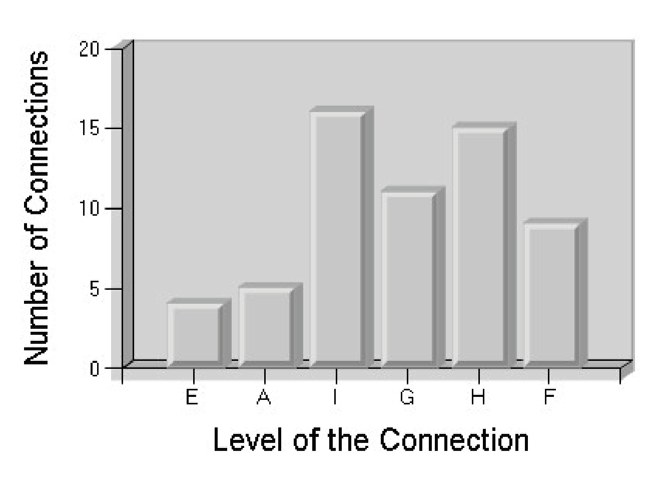

Fig. 4: SPARQL query of connections at different levels and integrated search result.

C. Finding Area Experts

To address a specific technical problem encountered at work, or assign a task, one may refer to area experts in the company. We present the following query to search for experts at some topic. This kind of queries are useful to speed up the solving process of a problem, or the completeness of a task. It can also eliminate employees frustration due to disproportionate or "irrelevant" task assignment.

Query 3: Get the employees who have a specific expertise level in a certain topic

```
PREFIX EOSK:<http://../Employee.rdf#>
SELECT ?EmployeeID
WHERE {
   ?EmployeeID EOSK:hasExpertise
?ExpertiseID.
   ?ExpertiseID EOSK:hasExpertiseLevel
EOSK:Expertise_Level_3.
   ?ExpertiseID EOSK:hasinTopic
EOSK:Topic_7.
}
```

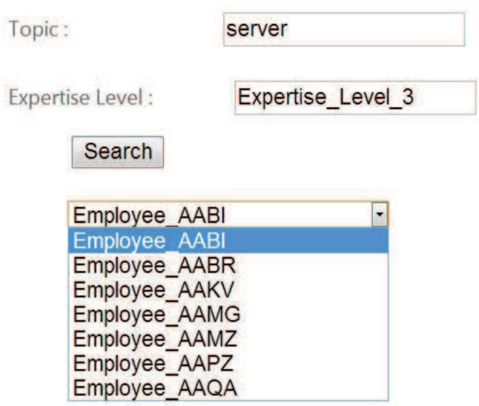

Fig. 5: SPARQL query of area experts and search result.

D. Trending Topics and Keywords

Discovery of trending topics is a typical application in social network analytics. In enterprise context, such macro analysis can prove to be more useful for identifying employees' trending interests, for instance, hot topics and trending employee users for each communication channel, department, and organizational position. In our ontology, "topics" are also used as specification of individual "expertise" and "interests". We can refer to the "keywords" in a "topic" as concrete descriptions of expertise areas and interests. Given a trending topic, the following query retrieves its most prominent keywords, in all contexts. This is direct result of performing topic modeling for text messages in enterprise microblogs.

Query 4: For a topic, get probability distribution of keywords in that topic

```
PREFIX EOSK:<http://../Employee.rdf#>
SELECT ?Topic_Keyword_No ?Keyword
?Probability
WHERE {
    EOSK:Topic_11 EOSK:hasKeyword
?Topic_Keyword_No.
   ?Topic_Keyword_No EOSK:hasKeyword
?Keyword.
   ?Topic_Keyword_No
EOSK:hasProbability ?Probability.
}
```

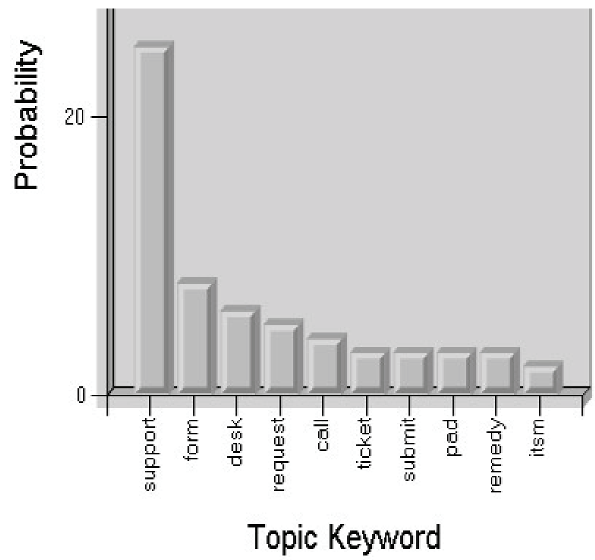

Fig. 6: SPARQL query of trending topic and search result.

VII. Conclusion and Future Work

In enterprises, it is fundamental to provide human resource database for attribute-based search. Common queries include finding a group of employees that fall in a specific range of ages, looking for someone at a specific position to replace another one, and identifying someone's supervisor. Ontology is used to represent the concepts and relations between them in modeling business processes. Existing methods can only support those simple queries based on single or multiple attribute filtering. However, they cannot satisfy complex semantic search such as expert finding, frequent contacts with common knowledge/intersts, and trending topics that are interested to employees. To address those problems, we leverage the usage of enterprise social media among employees. Specifically, we identify employees' expertise and interests by analyzing their informal communication at the workplace. We show how our employee ontology, enhanced with knowledge discovery from social networks can support complex semantic search in various corporate contexts. For future work, we plan to extend our ontology to recommendation systems that can personalize microblogging messages, and can automatically assign a team of employees to a new project.

VIII. ACKNOWLEDGEMENT

We would like to thank Gagandeep Singh, Vinodh Suresh and Swapnil Savdekar for the development of the instance generation tool and search interface.

REFERENCES

[1] G. Abowd, A. Dey, P. Brown, N. Davies, M. Smith, and P. Steggles. Towards a better understanding of context and context-awareness. In *Handheld and Ubiquitous Computing*, pages 304–307. Springer, 1999.

[2] L.A. Adamic, J. Zhang, E. Bakshy, and M.S. Ackerman. Knowledge sharing and yahoo answers: everyone knows something. In *Proceeding of the 17th international conference on World Wide Web*, pages 665–674. ACM, 2008.

[3] T. Bauer and D.B. Leake. Real time user context modeling for information retrieval agents. In *Proceedings of the tenth international conference on Information and knowledge management*, pages 568–570. ACM, 2001.

[4] D.M. Blei, A.Y. Ng, and M.I. Jordan. Latent dirichlet allocation. *The Journal of Machine Learning Research*, 3:993–1022, 2003.

[5] J. Chang and D.M. Blei. Relational topic models for document networks. In *Artificial Intelligence and Statistics*, 2009.

[6] J. DiMicco, D.R. Millen, W. Geyer, C. Dugan, B. Brownholtz, and M. Muller. Motivations for social networking at work. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, pages 711–720. ACM, 2008.

[7] K. Ehrlich, C.Y. Lin, and V. Griffiths-Fisher. Searching for experts in the enterprise: combining text and social network analysis. In *Proceedings of the 2007 international ACM conference on Supporting group work*, pages 117–126. ACM, 2007.

[8] G. Erétéo, F. Limpens, F. Gandon, O. Corby, M. Buffa, M. Leitzelman, P. Sander, et al. Semantic social network analysis: A concrete case. *Handbook of Research on Methods and Techniques for Studying Virtual Communities: Paradigms and Phenomena*, pages 122–156, 2011.

[9] J. Fink, A. Kobsa, and A. Nill. Adaptable and adaptive information access for all users, including the disabled and the elderly. In *International Conference UM97. Wien New York: Springer*, volume 171, page 173, 1997.

[10] S. Gauch, J. Chaffee, and A. Pretschner. Ontology-based personalized search and browsing. *Web Intelligence and Agent Systems*, 1(3/4):219–234, 2003.

[11] D. Godoy and A. Amandi. User profiling in personal information agents: a survey. *The Knowledge Engineering Review*, 20(04):329–361, 2005.

[12] M. Golemati, A. Katifori, C. Vassilakis, G. Lepouras, and C. Halatsis. Creating an ontology for the user profile: Method and applications. In *Proceedings of the First RCIS Conference*, pages 407–412, 2007.

[13] A. Kobsa. User modeling: Recent work, prospects and hazards. *HUMAN FACTORS IN INFORMATION TECHNOLOGY*, 10:111–111, 1993.

[14] T. Lappas, K. Liu, and E. Terzi. A survey of algorithms and systems for expert location in social networks. *Social Network Data Analytics*, pages 215–241, 2011.

[15] A. McCallum, X. Wang, and A. Corrada-Emmanuel. Topic and role discovery in social networks with experiments on enron and academic email. *Journal of Artificial Intelligence Research*, 30(1):249–272, 2007.

[16] P. Mika. Ontologies are us: A unified model of social networks and semantics. *The Semantic Web–ISWC 2005*, pages 522–536, 2005.

[17] D. Mimno and A. McCallum. Expertise modeling for matching papers with reviewers. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 500–509. ACM, 2007.

[18] R. Nallapati and W. Cohen. Link-plsa-lda: A new unsupervised model for topics and influence of blogs. In *International Conference for Weblogs and Social Media*, 2008.

[19] R.M. Nallapati, A. Ahmed, E.P. Xing, and W.W. Cohen. Joint latent topic models for text and citations. In *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 542–550. ACM, 2008.

[20] D. Raz, A. Juhola, J. Serrat, and A. Galis. *Fast and efficient context-aware services*. John Wiley & Sons, Ltd, 2006.

[21] E. Rich. Users are individuals: individualizing user models. *International journal of man-machine studies*, 18(3):199–214, 1983.

[22] Michal Rosen-Zvi, Thomas Griffiths, Mark Steyvers, and Padhraic Smyth. The author-topic model for authors and documents. In *Proceedings of the 20th conference on Uncertainty in artificial intelligence*, UAI '04, pages 487–494, Arlington, Virginia, United States, 2004. AUAI Press.

[23] J. Stan, E. Egyed-Zsigmond, A. Joly, P. Maret, et al. A user profile ontology for situation-aware social networking. 2008.

[24] M. Steyvers and T. Griffiths. Probabilistic topic models. *Handbook of latent semantic analysis*, 427(7):424–440, 2007.

[25] M.R. Tazari. A context-oriented rdf database. In *Proc. SWDB*, volume 2003, pages 63–78, 2003.

[26] M.R. Tazari, M. Grimm, and M. Finke. Modelling user context. In *Proceedings of 10th International Conference on Human-Computer Interaction*, 2003.

[27] J. Teevan, S.T. Dumais, and E. Horvitz. Personalizing search via automated analysis of interests and activities. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 449–456. ACM, 2005.

[28] J. Trajkova and S. Gauch. *Improving ontology-based user profiles*. PhD thesis, University of Kansas, Electrical Engineering and Computer Science, 2003.

[29] A. Von Hessling, T. Kleemann, and A. Sinner. Semantic user profiles and their applications in a mobile environment. *Artificial Intelligence in Mobile Systems 2004 (AIMS 2004)*, page 59, 2004.

[30] Q. Wang, H. Jin, and Y. Liu. Collaboration analytics: mining work patterns from collaboration activities. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, pages 1861–1864. ACM, 2010.

[31] X.H. Wang, D.Q. Zhang, T. Gu, and H.K. Pung. Ontology based context modeling and reasoning using owl. In *Pervasive Computing and Communications Workshops, 2004. Proceedings of the Second IEEE Annual Conference on*, pages 18–22. Ieee, 2004.

[32] 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*, pages 1955–1964. ACM, 2010.

[33] J. Zhang, M.S. Ackerman, and L. Adamic. Expertise networks in online communities: structure and algorithms. In *Proceedings of the 16th international conference on World Wide Web*, pages 221–230. ACM, 2007.