Using a Reputation Framework to Identify Community Leaders in Ontology Engineering
(Short Paper)

Christophe Debruyne and Niels Nijs

Semantics Technology and Applications Research Lab, Vrije Universiteit Brussel
firstname.lastname@vub.ac.be

Abstract. Ontology-engineering methods often prescribe roles and responsibilities. A problem is to identify who will play what role in an ontology-engineering project. We present a generic reputation framework for identifying leaders in a community of stakeholders. Reputation scores are based on analyzing the interactions users have with each other and a collaborative ontology-engineering platform. The framework was applied in an experiment involving 36 users that lasted for 8 weeks. We compared the results with the leaders identified by the participants via a survey and noticed an important overlap. Our results thus show the feasibility of identifying leaders in an ontology-engineering project based on reputation, which can then be assigned additional responsibilities.

Keywords: Ontology Engineering, Roles and Responsibilities.

1 Introduction

Ontology engineering is a social process and methods for building ontologies often prescribe roles and responsibilities. We argue that certain roles can be identified by analyzing one's interactions on a collaborative ontology-engineering platform. The role we will focus on is that of a "community leader", the person driving the ontology-engineering project, e.g., by guiding the discussions within a community of stakeholders. To this end, we identify the characteristics of a community leader and propose a reputation framework that assesses a person via sensors monitoring those characteristics.

The paper is organized as follows: Section 2 presents related work and the characteristics of a community leader, Section 3 proposes a reputation framework for a collaborative setting, Section 4 describes the reputation sensors that give scores to particular leader characteristics, Section 5 presents the results of an ontology-engineering experiment in which the framework was applied and Section 6 concludes the paper and presents future work.

2 Background and Related Work

Trust and reputation systems offer a mechanism to *"collect, distribute and aggregate feedbacks [from a community member's history]"* [16]. These feedbacks

R. Meersman et al. (Eds.): OTM 2013, LNCS 8185, pp. 677–684, 2013.

help others to make a better judgement and encourage trustworthy service or behavior in future interactions [13]. Different ways to use, calculate and represent reputation have been presented in literature for different goals: (i) increase the reliability and trust between agents [17], (ii) contribute to quality identification [10], (iii) improve contribution quality [13,12,10,16,15,3], (iv) filter and recommend content [13,11,10], (v) build or increase co-operation [9,4], (vi) offer improved & real-time business intelligence [5].

Few examples of assigning responsibilities based on trust and reputation exist. One example is Slashdot[1], where moderators are randomly selected to rate contributions based on the ratings from others on their own contributions. Slashdot offers more features to users with a good reputation during a certain period of time. This period is not in function of fixed time, but in function of amount of actions. In [20], influential users are being identified within microblogging services such as Twitter[2] by proposing TwitterRank, an extension of the PageRank[3] algorithm. An approach for identifying experts in a social network by analyzing the agent's local information and their network using graph-traversal was introduced in [22]. A graph representation is also used in [19], where social influence is modeled in their proposed influence propagation method.

In this paper, we aim to identify leaders by means of reputation scores. Leadership is similar to socially influencing group's members to improve collaboration between peers for achieving a goal [18]. Since our goal is to identify community leaders in a collaborative setting, we first need to know what characterizes such leaders. Defining these characteristics allows us to determine what needs to be observed and rated in a reputation framework. The following list of characteristics is based on [7,14,2,1]: (C1) Energy, passionate persistence & optimism, (C2) Goal-Driven, (C3) Build Trust, (C4) Willing to take risks, (C5) Pull and communicate with others, (C6) Work systematically, (C7a) Share knowledge, power and credit, (C7b) Work interdependently, and (C8) Understand others. Items C7a and C7b were considered as one in [1]. We choose to separate both aspects as we aim to measure these aspects separately.

3 Reputation Framework

Based on [13,16,15,6], we define a reputation framework as follows: *"a Reputation Framework is a general independent mediation system to facilitate trust and reputation within social platforms by rating the quality of the users' contribution and use these ratings to calculate a user's reputation score."* The core of a reputation framework is a *reputation computation engine* [13]. This engine will return the reputation score for a user. In this section, we describe such an engine. We define $\mathcal{A}$ as the set of all human agents and $\mathcal{P}$ as the set of all platforms. A platform is an abstract notion for a setting where agents interact with each other. A platform can be an information system (a forum), or a part

[1] `http://www.slashdot.org`
[2] `https://twitter.com/`
[3] `http://en.wikipedia.org/wiki/PageRank`

of an information system (a specific thread in a forum). The set of reputation results $\mathcal{R}$ is defined as $[0;100] \cup \{\oslash\}$, where $\oslash$ will be used when a (sub)result for a user does not exist. A reputation computation engine provides a mapping from a subset of agents $A \in 2^{\mathcal{A}}$ to elements of $\mathcal{R}$ by means of a set of platform configurations C. A platform configuration is a triple $\langle p, w_p, S\rangle$, where $p \in \mathcal{P}$ is a platform, w_p is the platform's weight and S a set sensor configurations. A sensor configuration is a pair $\langle s, w_s\rangle$ where s is a reputation sensor and w_s is the weight for that sensor on that platform. A reputation sensor is a function $s : \mathcal{P} \times \mathcal{A} \rightarrow \mathcal{R}$ returning a result for a given user on a given platform. The platform configuration result for user a and for platform configuration $\langle p, w_p, S\rangle$ is described as:

$$\rho_{\langle p,S\rangle}(a) = \begin{cases} \oslash & \text{if } |S'| = 0 \\ \dfrac{\sum_{\langle s,w_s\rangle \in S'} s(p,a) \times w_s}{\sum_{\langle s,w_s\rangle \in S'} w_s} & \text{if } |S'| > 0 \end{cases}$$

where $S' = \{\langle s, w_s\rangle | \langle s, w_s\rangle \in S \wedge s(p,a) \neq \oslash\}$

With the platform configuration result for a user described above, we can now describe the result of a set of platform configurations C for user a as:

$$\rho_C(a) = \begin{cases} \oslash & \text{if } |C'| = 0 \\ \dfrac{\sum_{\langle p,w_p,S\rangle \in C'} \rho_{\langle p,S\rangle}(a) \times w_p}{\sum_{\langle p,w_p,S\rangle \in C'} w_p} & \text{if } |C'| > 0 \end{cases}$$

where $C' = \{\langle p, w_p, S\rangle | \langle p, w_p, S\rangle \in C \wedge \rho_{\langle p,S\rangle}(a) \neq \oslash\}$

A user will be filtered when the value for each platform configuration for that user is $\oslash$, which means that this user had no monitored activity. Platform configuration for which **at most one** user has a value between $[0; 100]$ are also not taken into account. In the case of no users with such a value, the platform will not provide any information. In the case of only one user with such a value, we only have one user that is "active" with respect to the monitored activity and thus it would make no sense to find a leader for that platform configuration.

Then, for each platform configuration, we compute the z-scores for each value not equal to $\oslash$. Z-scores give an indication of the distance between the given value and the mean in a number of standard deviations and allow us to compare the z-score of a user across communities. We denote the z-score for platform configuration C_i and user a_j as $\zeta_{C_i}(a_j) = (\rho_{C_i}(a_j) - \mu_{C_i})/\sigma_{C_i}$ when $\rho_{C_i}(a_j) \neq \oslash$, otherwise $\oslash$ remains. According to the empirical rule, about 99.7 percent lie within 3 standard deviations from the mean. For each platform configuration, we rescale these z-scores such that they fit the range of [0;100] using this rule; for a platform configuration C_i and a user a_j, the z-scores are rescaled as follows:

$$\rho'_{C_i}(a_j) = \begin{cases} \oslash & \text{if } \rho_{C_i}(a_j) = \oslash \\ 0 & \text{if } \zeta_{c_i}(a_i) \leq -3\sigma_{C_i} \\ \frac{-50 \times (\zeta_{C_i}(a_j)+3)}{-3} & \text{if } \zeta_{c_i}(a_i) \in (-3\sigma_{C_i}; 0) \\ \frac{50 \times \zeta_{C_i}(a_j)}{3} + 50 & \text{if } \zeta_{c_i}(a_i) \in [0; 3\sigma_{C_i}) \\ 100 & \text{if } \zeta_{c_i}(a_i) \geq 3\sigma_{C_i} \end{cases}$$

Values above 50 indicate a better performance w.r.t. the mean on that platform configuration. For every user a_j, we compute the average of rescaled z-scores where values are not equal to $\oslash$. We then obtain a mapping between users and average rescaled z-scores. These average normalized z-scores provide us a ranking of leaders, a mapping of human agents $A' \subseteq A$ to $[0;100] \cup \{\oslash\}$ representing the Final User Reputation (FUR), computed for a user a_j as $\Sigma_{c_i \in C'} \rho'_{c_i}(a_j)/|C'|$ where $C' = \{c | c \in C \wedge \rho_c(a_j) \neq \oslash\}$, and returning $\oslash$ when $|C'| = 0$.

4 Reputation Sensors

This section introduces objective and subjective measurements by means of reputation sensors and show how are they related to the leader characteristics listed in Section 2.

Average Activity Rating. This *objective* reputation sensor returns a result for a particular agent, based on his activity within a discussion of the platform. The agent also gets credit for the intention of being active.

Reply Rating. Discussions drive the agreement processes. Even while agents are not obliged to participate, they can be involved by expressing their opinion on statements made by others (e.g., its relevance, constructiveness, etc.). A user can thus be given a score based on how others assess his contribution. As this sensor takes into account the opinion of others, we call this sensor *subjective.* An implementation of this sensor can be based on Slashdot's moderation system, which has been chosen as the system to implement for this paper.

Engagement Rating. This *objective* reputation sensor looks in what extent an agent allows other agents to be engaged within the discussions. In other words, does he allow enough time for others to participate and give an opinion?

Quality Assessment. People collaborate for a particular purpose and the outcome of this collaboration can be evaluated: e.g., to what extent is the artifact usable, correct, complete, etc. Assessing the usability of an artifact can be done via surveys, for instance, and is then a *subjective* sensor. In ontology engineering, one can for instance assess to what extent services are timely and correctly annotated with the evolving ontologies, which can be measured *objectively.* In our work, we will focus on the latter.

Interdependently Rating. We adopt the idea of [21] to build a social graph based on interactions, and apply it to the context of a collaborative setting with the agent's community as social network. Based on a user's participation, the user builds up his social graph with other community-members. This sensor looks at the graph of a user's graph. With this reputation sensor, we can measure the user's *communication characteristic objectively.*

Table 1 depicts how each sensor covers the characteristics of a community leader. There are two characteristics that are difficult to measure with above

Table 1. Leader Characteristics versus Reputation Sensors

	C1	**C2**	**C3**	**C4**	**C5**	**C6**	**C7.a**	**C7.b**	**C8**
Average activity rating	✓	✓			✓	✓			
Reply rating	✓		✓		✓				✓
Engagement rating		✓			✓				
Quality assessment	✓	✓				✓			
Interdependently rating					✓			✓	

mentioned reputation sensors. These characteristics are C4 and C7a, which are both difficult to assess since most take place as interaction outside a system. These aspects could be assessed via surveys, but would not scale well as these attributes can also evolve over time.

5 Experiment

The context of our experiment is a course on ontology engineering, part of the MSc in Computer Science curriculum of the Vrije Universiteit Brussel. In this experiment, 36 participants were divided into 10 groups. Each group had to build their own information system. The resulting conceptual schemas not only contained a view of reality that depended on the information chosen, but also on the perspective of each group. The whole group then had to build ontologies to enable semantic interoperability between their autonomously developed information systems as well as to annotate the database of an existing system. Ontology engineering started on March 10, 2013 and ended on May 10, 2013. Participants were aware that their interactions were taken into consideration to compute their reputation on the framework, but participants were also explicitly told they were not going to be evaluated on their reputation score. The method and tool the students used for ontology engineering is GOSPL[4] [8], in which (i) concepts are both described in terms of natural language definitions called glosses and using a formalism grounded in natural language, (ii) all ontology evolution operators have to be agreed before they are performed, and (iii) glosses drive the negotiation process. The group created multiple communities by separating concerns where each community had the goal to create an ontology for a particular (group of) concepts. Examples are the "Research Project Community" and "Publication Community" that created ontologies for respectively describing research projects and publications. Links between communities can be established by declaring synonyms, which is an agreement that two terms in two communities are deemed to refer to the same concept. In this experiment, we created a platform configuration for each such community. Each platform configuration was given the same weight. All reputation sensors were given weight 1 except for engagement rating, which was given weight 2 since discussing and voting is important for reaching agreements in GOSPL.

[4] Which stands for Grounding Ontologies with Social Processes and natural Language.

As no benchmark is available, we validate the result by comparing the ranking in the reputation framework with a survey sent to the participants. In this survey, we asked the participants up to three names of persons who they deemed to be leading the ontology project. The reputations scores for the participants are shown in Table 2. The column “Rep” represents the Final User Reputation in descending order. Each C_i depicts the platform configuration results for that user. 17 of the 36 participants replied to the survey. Table 3 depicts the number of times a certain participant was chosen as being a leader.

Table 2. Reputation scores of users

User	Rep.	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	c12	c13	c14	c15	c16	c17	c18	c19	c20	c21	c22	c23
x33	77			70														83					
x34	71		71																		73	69	
x4	63	67			71	64			63										62	50			
x14	62					66					88			32									
x17	62				64	59	39		88	72	74		42	64	75	65	45	66	53				
x22	61			23	70	60		67	76		63	52			66		64	55	72				
x3	59	70			58	58			58										61	50			59
x23	59	62			50				71										61	50			
x8	59			59														58					
x19	56			46														65					
x1	55			72														39					
x31	55			49														61					
x35	55		46																		57	61	
x20	54			58														50					
x30	52				57	72	69		59	32	46			64			37	23	64				
x11	52			32	70	68	63		54	46	64		35	69	52	27	75	46	49			34	50
x6	52		58								38				60								
x36	52			78					43		25				35			78					
x27	51			45														58					
x9	48	65							30														50
x21	48	39				62			41		50												
x10	48			47														48					
x2	47		57			46					58										40	60	23
x7	46	35				57			40		53			45									
x16	46	35				55			35		37			67									
x18	43	56			30	46			30										22				77
x5	43				34	26		35	47		44	52			62		30		60				
x24	41	14			28	17		66	58		61	24			47			40	59				
x29	41	47			43														33				
x12	41			48				32	29		39	71			29			40	40				
x32	40	59				47			44		23			25									41
x37	36		51			22					53										30	26	
x25	35				25	27	29		34		33		73	35	24	59		33	16				
x15	30			24														36					
x26	22																	22					
x13	17		17																				

Table 3. Number of votes per person in survey

Participant	x23	x17	x22	x11	x33	x36	x19	x8	x30	x14	x37	x31	x26
# votes	8	7	6	4	3	3	3	3	2	2	2	2	2

Some users are highly ranked even while they have no recognition by respondents. This can be explained by the fact that not all respondents participated in all communities, and thus their answers are based on the interactions they were involved with. We also note that the respondents did not participate in all communities; hence their answers are based on the interactions they have been involved in. When we then look at the FUR, a weighted average of all the user his rescaled z-scores, we observe that almost all indicated leaders from the survey have a FUR higher then 50. Which means that these users – in general – performed better than average in their respective communities.

6 Conclusions

In this paper, we defined and applied a reputation framework for identifying community leaders in an ontology-engineering project. The characteristics of a community leader were identified, for which different sensors were proposed. These sensors have been implemented as functions in our reputation framework. Our reputation framework was integrated in a collaborative ontology-engineering tool and conducted an experiment with 36 participants. As there is no benchmark available, we compared the results with a survey. We observed an important overlap between users identified as leaders in our framework and the persons regarded as leaders by the participants.

A limitation of the proposed framework is the absence of interactions *outside* the system. Face-to-face communication is difficult to capture and the analysis of – for instance – email communication, while more feasible, might rise privacy issues. One could draw inspiration from solutions such as Nepomuk[5] to collect this data. Future work also includes additional experiments and surveys to calibrate the weights for each sensor, and even for each platform. One can easily imagine ontology projects that are easier (e.g., light weight ontologies for annotation) than others (e.g., constraints needed for reasoning to support certain business processes). As future research we should also investigate whether the complexity of the ontology-project could have an impact on the scores.

Acknowledgements. We thank Prof. dr. Robert Meersman and Sven Van Laere for their valuable comments.

References

1. Archer, D., Cameron, A.: Collaborative Leadership: How to Suceed in an Interconnected World. Routledge (2009)
2. Carter, M.: The importance of collaborative leadership in achieving effective criminal justice outcomes. Center for Effective Public Policy (2006)
3. Casaló, L., Flavián, C., Guinalíu, M.: The role of perceived usability, reputation, satisfaction and consumer familiarity on the website loyalty formation process. Computers in Human Behavior 24(2), 325–345 (2008)
4. Cascella, R., Battiti, R.: Social networking and game theory to foster cooperation Input Paper at the 2nd ENISA Workshop on Authentication Interoperability Languages held at the ENISA/EEMA European eIdentity Conference (2007)

[5] http://nepomuk.kde.org/

5. Chang, E., Dillon, T., Hussain, F.: Trust and Reputation for Service-Oriented Environments: Technologies for Building Business Intelligence and Consumer Confidence. John Wiley & Sons (2005)
6. Chen, W., Zeng, Q., Wenyin, L., Hao, T.: A user reputation model for a user-interactive question answering system: Research articles. Concurr. Comput.: Pract. Exper. 19(15), 2091–2103 (2007)
7. Chrislip, D.: The Collaborative Leadership Fieldbook. J-B US non-Franchise Leadership. Wiley (2002)
8. Debruyne, C., Meersman, R.: GOSPL: A method and tool for fact-oriented hybrid ontology engineering. In: Morzy, T., Härder, T., Wrembel, R. (eds.) ADBIS 2012. LNCS, vol. 7503, pp. 153–166. Springer, Heidelberg (2012)
9. Du, P., Li, Y., Xu, W.: Cooperation evolution of indirect reciprocity by discrimination. In: Deng, X., Ye, Y. (eds.) WINE 2005. LNCS, vol. 3828, pp. 768–776. Springer, Heidelberg (2005)
10. Farmer, R., Glass, B.: Building Web Reputation Systems, 1st edn. Yahoo! Press, USA (2010)
11. Golbeck, J., Hendler, J.: Reputation network analysis for email filtering. In: Proceedings of the First Conference on Email and Anti-Spam, vol. 44, pp. 54–58 (2004)
12. Javanmardi, S., Ganjisaffar, Y., Lopes, C., Baldi, P.: User contribution and trust in wikipedia. In: CollaborateCom, pp. 1–6. IEEE (2009)
13. Jøsang, A., Ismail, R., Boyd, C.: A survey of trust and reputation systems for online service provision. Decision Support Systems 43(2), 618–644 (2007)
14. Linden, R.: The discipline of collaboration. Leader to Leader 2003(29), 41–47 (2003)
15. Noorian, Z., Ulieru, M.: The state of the art in trust and reputation systems: A framework for comparison. JTAER 5(2), 97–117 (2010)
16. Resnick, P., Kuwabara, K., Zeckhauser, R., Friedman, E.: Reputation systems. Commun. ACM 43(12), 45–48 (2000)
17. Sabater, J., Sierra, C.: Review on computational trust and reputation models. Artif. Intell. Rev. 24(1), 33–60 (2005)
18. Stogdill, R.: Handbook of leadership: a survey of theory and research. Free Press (1974)
19. Tang, J., Sun, J., Wang, C., Yang, Z.: Social influence analysis in large-scale networks. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 807–816. ACM (2009)
20. Weng, J., Lim, E.P., Jiang, J., He, Q.: Twitterrank: finding topic-sensitive influential twitterers. In: Proceedings of the Third ACM International Conference on Web Search and Data Mining, pp. 261–270. ACM (2010)
21. Wilson, C., Boe, B., Sala, A., Puttaswamy, K.P., Zhao, B.Y.: User interactions in social networks and their implications. In: Proceedings of the 4th ACM European Conference on Computer Systems, EuroSys 2009, pp. 205–218. ACM, New York (2009)
22. Zhang, J., Tang, J., Li, J.: Expert finding in a social network. In: Kotagiri, R., Radha Krishna, P., Mohania, M., Nantajeewarawat, E. (eds.) DASFAA 2007. LNCS, vol. 4443, pp. 1066–1069. Springer, Heidelberg (2007)