

Stakeholder Detection for Online Debates

Antoine Seilles

LIRMM

UMR 5506 - CC 477

161 rue Ada

34095 Montpellier France

Antoine.Seilles@lirmm.fr

Jean Sallantin

LIRMM

UMR 5506 - CC 477

161 rue Ada

34095 Montpellier France

Jean.Sallantin@lirmm.fr

Nancy Rodriguez

LIRMM

UMR 5506 - CC 477

161 rue Ada

34095 Montpellier France

Nancy.Rodriguez@lirmm.fr

ABSTRACT

This paper presents a work in progress about stakeholder detection for online debates. We propose an approach based on classical community detection methods applied to semantic social networks representation. We defend that new web2.0 tools should assist users to define semantic relations between users, groups and roles based on social interaction analysis. The main goal is to provide new mechanisms for moderation decreasing misunderstandings and highlighting unexpected behaviors.

Categories and Subject Descriptors

H.5.3 [Group and Organization Interfaces]: Organizational design, Web-based interaction.

General Terms

Experimentation.

Keywords

E-democracy, E-communities, Stakeholder detection, Community detection, Opinion, Annotation, Role, Relation

1.INTRODUCTION

The web offers new hopes for e-democracy, providing many new ways of interacting between citizens and civil servants. One of the key feature of democracy is the possibility to debate. Websites for online debates use classical web2.0 tools for enabling discussions: forums, blogs, instant messenger, etc. However, these tools have not been designed for debates and have scalability limitations. We argue that two major aspects have to be taken into account for online debates: the possibility to express opinions and stakeholders. Within a large scale debate, one needs to know the main tendencies: what are the alternatives options and which group supports the various options. Our approach is to use semantic web technologies to represent users interactions and build new methods of communities and tendencies detection.

This paper starts with a reminder of the principal hopes provided by e-democracy and the current limitations of online debates. Then, we present an alternative solution to classical opinion mining approaches. Our solution is based on semi-structured annotations. We present a model to unravel a web user's identity and we present how this model matches with existing RDF ontologies. Then, we present Stakeholder detection as a

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

IIWAS2010, 8-10 November, 2010, Paris, France.

Copyright 2010 ACM 978-1-4503-0421-4/10/11...\$10.00.

particular case of community detection. We discuss the notion of relationship and classical graph-based representations for social networks. Finally, we present our experimental approach to support stakeholder detection and organization.

2.E-DEMOCRACY AND ONLINE DEBATES

The main goals of e-democracy are to facilitate open exchanges of controversial political ideas, to increase citizens participation and to reduce organizational costs of consultation [22] [25] [5]. Web tools for e-democracy facilitate the expression of opinions due to anonymity and due to accessibility from everywhere at anytime. This ease of access is also a factor to decrease organizational costs. Nowadays, most of the web tools used for e-democracy are discussion tools (such as blogs or forums) dedicated to e-consultation. These tools have a cost of moderation and animation [4]. A reason for this moderation and animation cost is the lack of an executive summary. Web applications that provide debate summaries need to process data to extract opinions and categorize the statements. They are not linked to discussion tools. There is no available tool to structure an argumentation (the best feature currently is the possibility to type a statement as pro or con using modalities such as "I like" or "I agree/disagree"). This lack of structure and summary leads to huge costs in data processing of large scale debates.

Another point to highlight is that democracy is not only made of debates. Debate in democracy is only one step of a collaborative process to decide [25]. A dedicated tool for democratic discussions should be inter-operable with other tools (online or not), like calendars, pole applications, social network application, etc.

We have adopted the web semantic technology of RDF (Resource Description Framework) to represent debates data (opinions and argumentative structures) and to facilitate interoperability. RDF allows to represent relations between subject and object and is a standard for data exchanges.

3.ANNOTATION EXPRESSING OPINIONS

3.1 Limitation in opinion mining

[2] defines an opinion as a discursive act in which the speaker "models" explicitly or implicitly the object of his statement with the dimensions "possible/impossible", "desirable/undesirable", "beautiful/ugly", "pleasant/unpleasant", etc. Opinion mining is one application of the natural language processing domain. The main assumption in [12] is that an opinion can be detected by the adjectives used. Two kind of methods exist to detect opinions: using a corpus of opinions done by experts [21] or using adjectives repositories such as WordNet [16] [14]. Results will largely differ depending on the corpus used. One adjective may express a different opinion depending on the context [28]. All these methods are not dynamic and context related. We

suggest to use semi-structured annotations as a dynamic alternative to opinion mining.

3.2 Semi-structured annotation

[26] defines annotations as any new information added to web data. [15] splits annotations in two categories: semantic annotations are meta-data added to a document, discursive annotations are comments attached to a document and linked together by a discursive structure. Both annotations can be associated to a full document or only to a part of a document.

Semantic annotations are the kind of meta-data RDF can express. Most of the time, these semantic annotations are produced by professionals specialized in knowledge representation. A discursive annotation is similar to a forum post except that the discussion can be represented by a graph instead of a thread (posts can have multiple replies and are ordered by reply status). One interesting feature of discursive annotations is the possibility to highlight the part of a post to which a comment replies. This feature is badly implemented in most of the current forums by the "quote" functionality. This property allows users to highlight the core statements.

[20] offers solutions for users to produce meta-data while they enter a post in a forum or an article in a wiki. Using the same methodology, we present a solution for producing both semantic and discursive annotations. Figure 1 shows one of our annotation forms. It contains a choice of "emoticons" to express a mood, a choice of labels to express an opinion and text-areas to add comments. Once validated, these comments are stored using RDF vocabularies: SIOC, DublinCore, RSS1.0, and Annotea annotation Schema. Currently, these vocabularies don't allow to represent the core information to enhance social network analysis: opinions, moods and comments status. [3] offers an ontology to express comments status. The AIF ontology allows to represent argument networks. The NiceTag ontology [17] is designed to represent the action of tagging. It's possible to represent the fact that a tag expresses an opinion. Figure 2 shows a limited representation of an annotation. We have not yet included ontologies such as AIF or NiceTag.

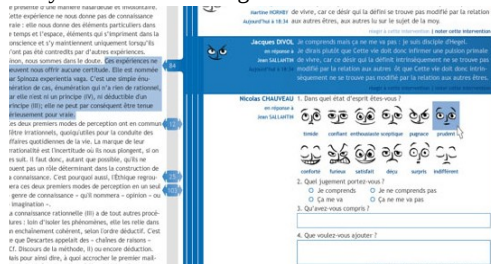


Figure 1. An annotation form

```
<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:sioc="http://rdfs.org/sioc/ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:annotea="http://www.w3.org/2000/10/annotation-ns">
  <annotea:Annotation rdf:about="http://www.natoine.fr/annotation/?p=479">
    <annotea:annotates rdf:resource="http://www.jeanSallantin.fr/homepage"/>
    <annotea:body>
      <sioc:Post rdf:about="http://www.natoine.fr/wordpress/?p=401"
        dc:title="IPC ou communication inter portlet 2. JSP286"
        rdfs:seeAlso rdf:resource="http://www.natoine.fr/sioc"/>
      </sioc:Post>
    </annotea:body>
    <annotea:context>
      http://www.jeanSallantin.fr/homepage#pointer(id="Main")/p[2]
    </annotea:context>
  </annotea:Annotation>
</rdf:RDF>
```

Figure 2. XML/RDF export of a discursive and semantic annotation

4. DIGITAL IDENTITY REPRESENTATION

[9] presents a model to unravel users' identity. The digital identity is a representation of the user's declared profile, activity and statistics. The first component of digital identity is the declarative identity constituted of every data the user give to the system about himself. Typically, when a user subscribes to a new application he is invited to fill a form containing personal details (address, age, gender, interests, friends, etc.). Declarative identity data can be represented by the FOAF ontology. The FOAF ontology is dedicated to represent personal data and relationships. The second component of digital identity is the acting identity constituted of all interactions the user has through the system. For example, when he modifies his status or when he sends a message. The SIOC ontology offers a partial representation of the acting identity. The last component of digital identity is the calculated identity. It is constituted of numbers calculated by the system and concerning the user (for example, statistics on his browsing or chatting activities). Results of classical Social Network Analysis methods are parts of the computed identity. The ontology SemSNA [7] is dedicated to represent results of such methods. Figure 3 summarizes a matching between digital identity and ontologies.

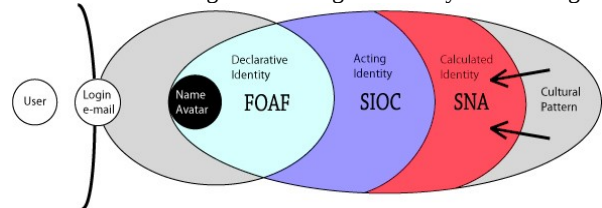


Figure 3. Ontologies to Represent Digital Identity

We will defend in the next part of this paper that, by using results of SNA methods applied to the acting identity data, it becomes possible to compute and offer a structured representation of social networks. By structured representation of social networks, we mean a social network where individuals are represented "linked together" by labeled relationships. They also are associated to the roles they play in the network (leader, troll, etc.). In other words, we say that it's possible to deduce a FOAF representation of someone's social network using SNA methods applied to a SIOC representation.

5. STAKEHOLDER MANAGEMENT

5.1 Stakeholder detection

Stakeholders are people sharing interests and opinions. Most of the time, stakeholder groups are already organized and act as lobbying groups during debates. We are interested in detection of unorganized and unidentified stakeholder groups. According to [1], these groups deserve more attention than already existing stakeholder groups. They stress new doubts about issues of the debate and are weaker than existing groups. The dynamic of a debate is correlated to the dynamic of stakeholder groups. A new doubt creates new subjects of debate and knowledge emerges of each debate. New knowledge creates new doubts. But stakeholder detection is not only a simple community detection problem. Community detection is based on representing social networks by mono-relational graphs. Stakeholder detection is based on expressing opinions. We share the same conclusion as [6], social network analysis need a social network representation concerned with semantics not only with syntax as it is the case in classical graph-based representations.

5.2 Community detection

In the community detection domain, the definition of a community is only linked to sociometry and graph theory and not to a true social theory. [8] presents a review of community detection methods and defines communities as being “clusters or modules, groups of vertices which probably share common properties and/or play similar roles within the graph”. In such methods, only the amount of relations and of nodes in the graph is concerned. The roles of individuals, the different natures and intensities of relationships are not taken into account. Indeed, they are presented as an explanation and supposed unknown. Community detection lays on mathematical properties of specific graphs patterns [23]: components (isolated connected subgraphs), cycles (paths where starting and ending vertices are the same) and cliques (each vertex is linked to all others). Referring to Granovetter's theory of strength of weak ties [10], such methods can successfully detect groups of individuals strongly tied each others and similar. Granovetter in [11] revisits its own theory through a review of empirical studies. Most of his assumptions seem to be validated, but depending on the context, the interpretations of the assumptions differ. [10] proposes that every relationship can be quantified measuring an intensity or a strength. The definition given for the strength of a tie is the following : “a combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie”. A group is characterized by the fact that individuals are tied through strong ties. Two individuals strongly linked must share many common personal characteristics. This tends to define a group as a set of similar individuals and implies that a bridge between two groups might be a weak tie. Figure 4 summarizes the impact of strong and weak ties in community detection. Empirical studies [11] show that the definition of the strength of ties depends on the context and the nature of the tie. Two persons with a strong tie for business are not necessary strong friends. Indeed, this idea may refer to the notion of *Netdom* developed by White [27]. White defends that a network is composed of different sub-networks depending on domains. Two individuals may be differently linked depending on the domain (business, family, politic, etc.).

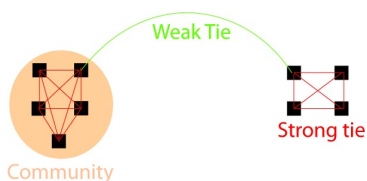


Fig 4. Strong and Weak ties in community detection

5.3 Web interactions, Communities and Roles

The web provides many different ways to interact between users. These interactions are data for social networks analysis. Some platforms such as Facebook are dedicated to build social network. Another approach is to extract social network from existing interactions (email [24], IRC [18], etc.). Referring to the Nadel's theory of social structure [19], we adopt the strategy of extracting a social network from a set of interactions between users. Nadel defines a relationship as a set of social interactions between two individuals. Using tagging and annotation activity, it becomes possible to detect two users sharing the same interests and supporting a same claim. This approach is similar to recommendation systems based on personal activities [13]. Most of recommendation algorithms are based on similarity of

profile or activities history. Referring to Granovetter's theory, classical community detection and recommendations approach based on similarity are effective and constitute the core of a stakeholder community. However, similarity doesn't provide a solution to organize a group. A complex organization is not made only of similar individuals. Most of the time, an organization is the coordination of complementary individuals. Furthermore, the lack of heterogeneity in a group due to similarity has a perverse effect : polarization and radicalization of opinions. This effect is one hypothesis of Granovetter's theory and is one of the effects feared by Price [22] about online discussions. Therefore, we suggest to extract a social network not only on similarity but also on opposition and weak ties.

Nadel defends that a society is structured by roles over relationships. An individual playing a role is supposed being aware of the role he plays. A role defines a behavior through possessions and skills and through relationships between the individual playing the role and others. Because a relationship is defined by a set of interactions, an individual playing a role is supposed to act a specific way in each situation. Because the definition of a role is supposed to be known by everybody, an individual playing a role is exposed to social sanctions if he doesn't behave correctly.

We base our approach on representation and storing of rich social interactions. We can detect groups sharing a same opinion about a subject querying RDF representation. We can assist users to define their relationships and their roles revealing patterns of users behaviors. We try to create a recommendation system based on complementarity, not only on similarity. The next part of this paper presents an experimental approach and our expected results.

6. EXPERIMENTAL APPROACH

Our forthcoming experiment concerns users acceptability of roles and labeled relationships. Do users correctly use semi-structured annotations? Do semi-structured annotations provide rich interactions we can exploit to analyse the network of users?

Two groups of users have to discuss a given subject using semi-structured annotations. Each group is a set of individuals already knowing each other. Users of the first group are identified by first name and last name. Users of the second group are identified by anonymous pseudonyms. The system will log all the information concerning interaction between users: the use of a same tag, replies support or opposition, etc. After a determined duration, the system will give a summary of interactions between users and will ask them to identify relationships between them.

We expect the following results: Patterns should emerge demonstrating the existence of different relationships between users and the redundancy of these relationships. Patterns should emerge between users linked by the same type of relationships. It validates the existence of classes or categories of users which is a necessary condition for existence of roles. Users already know each other and already have interactions and relationships outside of the system. Therefore, users identified by last and first names should be easily aware of their relationships. The comparison between anonymous and non-anonymous users should stress that knowing each other identity has an impact on interactivity. Indeed, we expect that anonymous users will express more opinions.

7.CONCLUSION

We presented a context of development of new web 2.0 applications dedicated to online debates. We briefly discussed the need for web semantic representation. We discussed the need of new social network analysis based on labeled relationships. Applications offer more and more way to interact but offer poor vocabularies to define relationships between users. We suggest an approach based on Nadel's theory dedicated to organization of social networks. We argue that a structured social network, composed of individuals playing roles and linked by labeled relationships, represented by the FOAF ontology can be deduced using SNA methods applied to a representation of social activities or interactions such as the SIOC ontology.

8.ACKNOWLEDGMENTS

This research was sponsored by the Intermed-ANR project. Thanks to Martine Hornby for her technical support.

9.REFERENCES

- [1] Callon, M., Barthe, Y. and Lascoumes, P. 2002 Qu'en pensent les citoyens ? In magazine *Sciences humaines*, 124, Rubrique Société du risque : fantasmes et réalité.
- [2] Chabrol C., Bromberg M. 1999. *Préalable à une classification des actes de parole*, Psychologie Française, 44-4, 291-306, Paris.
- [3] Chesñevar, C., McGinnis, J., Modgil, S., Rahwan, I., Reed, C., Simari, G., South, M., Vreeswijk, G., and Willmott, S. 2006. Towards an argument interchange format. *Knowl. Eng. Rev.* 21, 4 (Dec. 2006), 293-316.
- [4] Confino, F. 2009. *Blogs territoriaux, réseaux sociaux et nouveaux enjeux du web2.0 pour les collectivités*. Territorial Editions. Dossiers d'expert 625.
- [5] Desquinabo, N., Ferrand, N. and Marlier, J. 2010. Stakeholder e-Participation in Local Planning: The Camargue Park Case. In A.B. Sideridis & Ch.Z. Patrikakis, (Eds.) *Next Generation Society Technological and Legal Issues*, eDemocracy 2009, LNICST, Vol. 26, pp. 128-137
- [6] Ereteo, G. , Buffa, M., Gandon, F., Grohan, P., Leitzelman, M. and Sander, P. 2008 A State of the Art on Social Network Analysis and its Applications on a Semantic Web, SDoW2008 (Social Data on the Web), workshop at the 7th International Semantic Web.
- [7] Ereteo, G., Gandon, F., Buffa, M. and Corby, O. 2009 Semantic Social Network Analysis, In Proc. Web Science WebSci'09, Athens, Greece.
- [8] Fortunato, S. 2010. Community detection in graphs, *Physics Reports*, 486 (3-5) pp 75-174.
- [9] Georges, F. 2009 Représentation de soi et identité numérique, *Réseaux 2*, vol 154, pp 165-193.
- [10] Granovetter, M.S. 1973. The Strength of Weak Ties, *American Journal of Sociology* 78(6): 1360-1380.
- [11] Granovetter, M.S. 1983. The Strength of Weak Ties : A Network Theory Revisited, *Sociological Theory*, Volume 1, 201-233.
- [12] Harb A., Plantié M., Dray G., Roche M., Troussset F., Poncelet P. 2008. Web opinion mining: how to extract opinions from blogs?, *CSTST '08: Proceedings of the 5th international conference on Soft computing as transdisciplinary science and technology*, Cergy-Pontoise, France, ACM.
- [13] Iskolde, A. 2007. The Art, Science and Business of Recommendation Engines, http://www.readwriteweb.com/archives/recommendation_engines.php
- [14] Liu B. 2007. Tutorial on sentiment analysis, based on Chapter 11 of the book *Web Data Mining – Exploring Hyperlinks, Contents and Usage Data* <http://www.cs.uic.edu/~liub>
- [15] Lortal G. 2006 Médiatiser l'annotation pour une herméneutique numérique : AnT&CoW, Thèse de doctorat, Université de Technologie de Troyes.
- [16] Miller, G. A. 1995. WordNet: a lexical database for English. *Commun. ACM* 38, 11 (Nov. 1995), 39-41.
- [17] Monnin, A., Limpens, F., Laniado, G. and Gandon, F. 2010 Speech acts meets tagging : NiceTag ontology. *5th AIS SigPrag International Pragmatic Web Conference Track (ICPW 2010) at the International Conference on Semantic Systems (i-Semantics 2010)*, September 1-3 Graz, Austria.
- [18] Mutton, P. 2004. Inferring and Visualizing Social Networks on Internet Relay Chat. In *Proceedings of the information Visualisation, Eighth international Conference* (July 14-16, 2004). IEEE Computer Society, Washington, DC, 35-43.
- [19] Nadel, S.F. 1957. *The Theory of Social Structure*, London, Cohen et West.
- [20] Passant A. 2009. Technologies du Web Sémantique pour l'Entreprise 2.0, Thèse de doctorat, Université Paris IV-Sorbonne.
- [21] Plantié M., Roche M., Dray G., Poncelet P. 2008. Is a Voting Approach Accurate for Opinion Mining?, *DaWaK'08: Proceedings of the 10th international conference on Data Warehousing and Knowledge Discovery*, Turin, Italy, Springer-Verlag.
- [22] Price, V. 2006. Citizen Deliberating Online : Theory and some Evidence. In T.Davies et B.S.Noveck (eds.), *Online Deliberation : Research and Practice*, Chicago University Press.
- [23] Scott, J. 2000. *Social Network Analysis : A Handbook*, SAGE publications, London UK.
- [24] Tyler, J.R., Wilkinson, D.M. and Huberman, B.A. 2003. Email as spectroscopy : automated discovery of community structure within organizations. In *Procedures C&T 2003*.
- [25] Vedel, T. 2003. L'idée de démocratie électronique : origines, visions, questions. In Pascal Perrineau (dir.), *Le désenchantement démocratique*, La Tour d'Aigues, Éditions de l'Aube.
- [26] Veronis J. 2000 Annotation automatique de corpus : panorama et état de la technique. In J.-M.Pierrel (dir) *Ingénierie des langues*, Paris: Editions Hermès, p111-129.
- [27] White, H.C. 2008. *Identity and Control*, second edition, Princeton.
- [28] Wilson, T., Wiebe, J., and Hoffmann, P. 2009. Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Comput. Linguist.* 35, 3 (Sep. 2009), 399-433.