Towards a Framework for Social Semiotic Mining

Eirini Giannakidou
Aristotle University of Thessaloniki
54124 Thessaloniki, Greece
+30 2310 991865
eirgiann@csd.auth.gr

Athena Vakali
Aristotle University of Thessaloniki
54124 Thessaloniki, Greece
+30 2310 998415
avakali@csd.auth.gr

Nikolaos Mavridis
NCSR Demokritos
15310, Agia Paraskeyi, Greece
+30 210 6503153
nmavridis@iit.demokritos.gr

ABSTRACT
Although the theory of semiotics arguably has ancient beginnings and came to the forefront with the seminal work of Pierce in the 20th century, and despite the growth of social media and the direct relevance of semiotics, no framework has so far been provided, which not only enables the re-examination of social content and tagging under the light of semiotics, but can also be used to analyze data mining and clustering algorithms utilized on social data. We provide the motivation and the outline of such a framework in the paper, and demonstrate how it can be applied not only in order to analyze specific algorithms, but also in order to structure the general space of potential algorithms for clustering data derived from social media.

Categories and Subject Descriptors
D.2.8 [**Information Systems**]: Database applications – *data mining,* J.4 [**Computer Applications**]: Social & behavioral science – *sociology*

General Terms
Algorithms, Design, Theory

Keywords
Social Media, Data Mining, Semiotics

1. INTRODUCTION
Social Networking Sites (SNS) and Online Social Media are central features of the Web 2.0 revolution which is well underway in the second decade of the 21st century. This has created a lot of interest towards the automated or semi-automated content analysis of such sources, in order to be able to extract knowledge, also taken into account the advantages of the wisdom of crowds [1], and to identify patterns. Thus, data mining techniques, such as clustering, have recently often been used on social media, towards a multitude of applications, including automated recommender systems [2], policy planning for local governments [3], organizational development [4], market research etc.

However, when applying such techniques to social media data, there seems to be a lack of a clear theoretical modeling of the underlying structures: how can one connect tags with objects with concepts, across users and time? As we shall illustrate, towards that goal, existing theories of Semiotics, can inform us for creating a theoretical framework for social data mining. Semiotics as a field deals exactly with the relation between signs, concepts, and referents; and thus, fruitful analogies with the social media will be derived.

In this paper, starting from the act of tagging (annotation), which constitutes one of the core processes in Social Web, where users assign arbitrary labels (tags) on digital resources, in order to describe or classify them, we see the three-fold relation that arises between signs / tags, referents / digital objects and concepts / representations. This relationship exists in all semiotic systems and described through the semiotic triangle, which will be described in detail below. On this triangle and utilizing several aspects of it such as similarity spaces arising within vertices, transformations across vertices, and second-level similarities, a clear theoretical framework for the problems dealt with in social media analysis will be provided. The rest of the paper is organized as follows. In the next section, we will present some basic theory about signs and their study from the field of Semiotics. Then, Social Semiotics are discussed and the idea of modeling social media resources as social semiotic resources is introduced. To support this idea, a mining process in social media is modeled through an extension of the language of social semiotics, in Section 4, where the generalized framework is presented, and applications are illustrated. Then, in section 5, a discussion is given, also including proposed future work. Finally, in section 6, we provide a concluding section.

2. SIGNS AND SEMIOTICS
Humans create concepts in their mind from an early age, in order to explain and classify the stimuli received from their environment. On a more collective basis, all contemporary cultures use a plurality of signs, whose meaning is assumed to be known to a significant portion of their population. These signs can be in the form of words, images, sounds, acts or objects. One way to classify signs is using two categories: natural signs and conventional signs. Conventional signs do not have intrinsic meaning on their own; they acquire meaning within a specific cultural context and only for a specific group of interpreters. "*Nothing is a sign, unless it is interpreted as a sign*" says Pierce [5]. Anything can become a sign as long as someone interprets it as signifying something, i.e. it refers to or stands for something else, beyond itself. In reality, a big number of entities are interpreted as signs, to a large degree subconsciously, through automatic associations with constructed systems of convention,

WIMS'14, June 2-4, 2014, Thessaloniki, Greece

ACM 978-1-4503-2538-7/14/06…$15.00.
http://dx.doi.org/10.1145/2611040.2611087

with which the receiver is familiar. The receiver often does not have concious reflection of that.

Semiotics as a field is occuppied with the study of signs and the relation to their referents and their interpreters. It is noteworthy that beyond conventional signs, as mentioned above, there are also natural signs: for example smoke can be construed as a natural sign of fire, given that when we observe smoke, we can infer the existence of fire (that has produced it). In this case, the relationship between a sign and its referent arises out of the causality of nature. In contrast, in the case of conventional (and not natural) signs this relation arises out of a social convention. It is worth noting that in ancient greek the word *σύμβολο* (that stands for sign) etymologically arises out of the composition of the constituents συν (together) and βάλλω (shoot). This is related to the fact that in a natural sign, the sign (smoke) and its referent (fire) are "shot" in close temporal proximity, and, thus, the observation of an overt perceptible entity (smoke) can support the hypothesis of the existence of another covered entity (fire).

The Semiotics can provide interesting theoretical frameworks for modeling concepts and processes related to technologies of the Social Web. As we will see below, the labels (tags) that are given by users can be considered as signs which refer to the digital objects (resources) residing in the web, and, at the same time, to concepts residing in the mental space of users. Thus, the basic apparatus of Semiotics, such as the semiotic triangle, which we will present below, can form the basis for a clear framework, through which we can theorize and analyze social media and data mining processes on them, as we shall see.

3. SOCIAL MEDIA AS SEMIOTIC RESOURCES

According to [7], Semiotic as a science searches for and studies the meaning and usage of signs. The *semiotic triangle,* also known as the *triangle of meaning*, is a depiction of the way that signs are related on the one hand with the objects that they refer to, and on the other hand with the concepts to which they correspond (see Figure 1). This specific triangle was published in 1923 by [6] , although the basic idea goes back at least to the 4th century BC in the Peri Hermineias (De Interpretazione) of Aristotle [8]. Aristotle differentiated between objects, the words that refer to them and their corresponding experiences of the sole (psyche). An alternative, extended version of the triangle is the one that is shown in Figure 2.

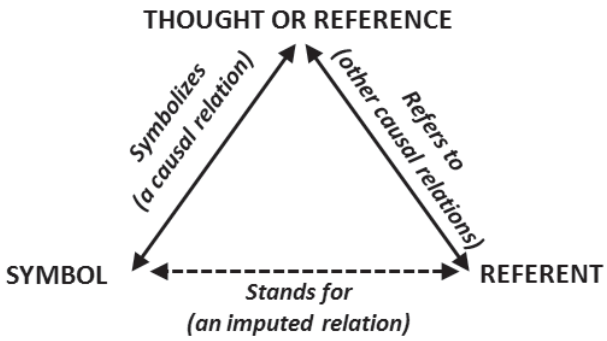

Figure 1: Semiotic Triangle [6].

A more complete collection of different variations of the semiotic triangle is presented in Figure 3 (presented in [9]), which includes not only the proposals of Ogden and Richards [6], but also of Pierce [5], Ullman [10], Stephen Harnad [11], and Vogt [12].

Social Semiotics is the branch of Semiotics that researches human practices of *semiosis*, i.e. of the connection of signs with concepts and objects under specific social and cultural conditions, and which tries to explain the attribution of meaning as a social practice [13]. More specifically, in social semiotics ways are examined through which different media of a society are used, in order to express the meaning of a concept. To these, obvious methods of communication are included, such as language, gestures, images and music together with less obvious, such as food, clothing and objects of everyday, which are however carriers of cultural value and meaning. Social semiotics also examine semiotic practices across differing cultures and communities. The set of actions and constructions which are utilized for the communication between humans are called *semiotic resources* and are being produced either through human physiology (for example, with our muscles we create expressions on our face or gestures, with our vocal chords we create sounds) or with the usage of technological means (for example, with pen and paper, with the software and hardware of computers, etc.)

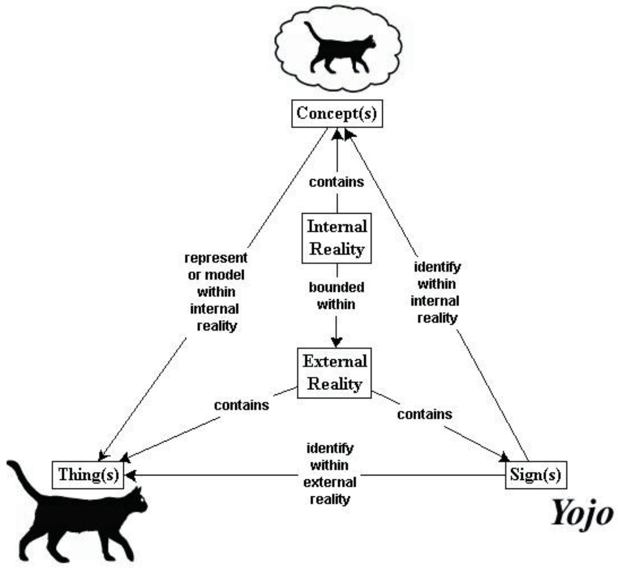

Figure 2: Extended version of semiotic triangle [14].

In our days, people, often, communicate with other people using a virtual environment in a computer through social media. As we mentioned before, in the Social Web millions of people interact via weblogs, collaborate in wikis, play games together, upload and describe digital content and build social networks between themselves. This new domain of communication introduces a new dimension to the field of Social Semiotics. Every activity from the ones mentioned before is tagged with labels from users. In Social Semiotics every sign is connected to a concept and an object, as depicted in the semiotic triangle (see Figure 1). Correspondingly, in social media every tag is connected to a digital object and to a concept that resides in the mind of the user. Thus, through the examination of this analogy, we propose the transfer of the semiotic triangle to the Social Web, as depicted in Figure 4.

Following this framework, in order to detect wider relations among tags, concepts and digital resources, we examine multiple semiotic triangles that arise from different users in different temporal instances, aiming towards the widest possible coverage and objectivity in our conclusions. Indeed, by studying multiple triangles and merging <TAG> vertices, a tag space arises, where the links between tags are based on any kind of relationship between them, such as co-occurrence or semantic affinity. If we,

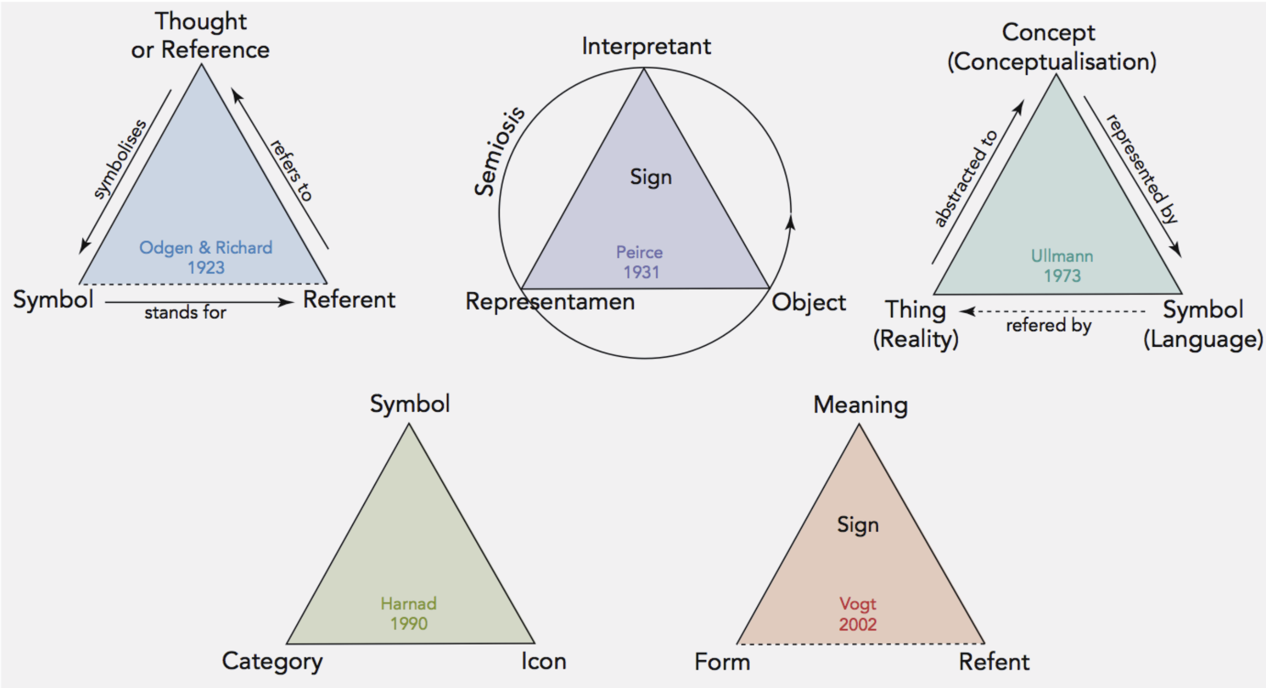

Figure 3: Variations of the Semiotic Triangle, according to [5],[6],[10],[11],[12].

then, define an appropriate similarity measure, a similarity space arises, metric or non-metric. Analogously to the similarity space that arises in the <TAG> vertex of the semiotic triangle, one can define a similarity space for the <RESOURCE> vertex, as well. In this space, in which the resources may be digital objects, such as photos, videos and so on, distances can be defined based on their low-level features. Finally, let us move on to the third vertex of the semiotic triangle <THOUGHTS OR REFERENCE>. Here, the entities could, for example, be: i) Concepts as represented in the mind of a user, ii) Concepts as represented in a semantic network (like WordNet [15]), iii) ideal forms, which are independent of the observer, such as the Platonic Forms [16]. In this third vertex, one can create a similarity space, too. For example, by defining distance of two concepts through the minimum number of edges needing to be traversed within WordNet to "go" from one concept to the other.

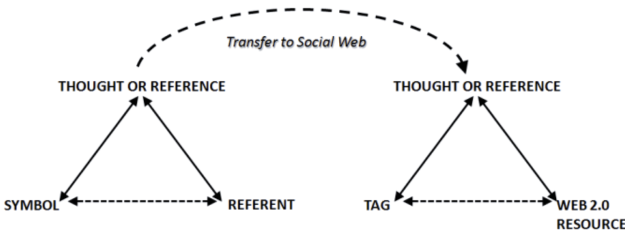

Figure 4: Transfer of semiotic triangle to social media.

Thus, having defined these spaces (one per each triangle vertex), we can search for generalized relationships between entities, in one of the following ways (depicted graphically in Figure 4):

- **Examination of first level relationships (taking into account entities in one vertex only, Figure 5 α)**

  Relations among entities in the same vertex (i.e. tags, digital objects or concepts) directly estimated from the corresponding similarity spaces, such as the ones we talked about in the preceding paragraphs.

- **Examination of second level relationships (taking into account entities in two vertices, Figure 5 β)**

  On the semiotic triangle, each vertex is connected to any other vertex through an edge. For example, tags are connected to digital objects that were tagged with them. Thus, when one wants to calculate the relation between two tags, he could (instead of remaining within the tag vertex only) jump to the object vertex too, i.e. the similarity between any tags could be calculated through the similarity of the objects that they correspond to (a second level relation, involving not only the tag vertex but also the object vertex)

- **Extension of the semiotic triangle across multiple axis (for example across the temporal axis or the user axis, Figure 5 γ and δ)**

  Given the dynamic and collective nature of Social Web, a very interesting kind of information arises when tracking relationships between entities across a timeline or across various user communities. Following the proposed framework, we can track the evolution of such relationships, by examining different snapshots of triangles along the axis, we are interested in.

4. SOCIAL MEDIA CLUSTERING

Here, we will present the problem of clustering of social data and analyze it within the framework of social semiotic mining.

Problem [Social data clustering problem]: Determination of meaningful clusters within social data through the examination of relations in multiple spaces. ■

As an illustration, here we will describe example cases, using: i) tags only, ii) tags and content, and iii) tags, content, users and time. Before we examine examples, we shall start by introducing a general framework, specific cases of which will be the examples that we will provide.

4.1 Generalized Social Clustering Framework

Our generalized framework consists of a number of steps:

Step 1) Chose **type of clustering**: One-way clustering (L1), Co-clustering (L2)

Step 2) Determine which **subset of vertices** U from V=(T,C,R) will be used for the distance function for the clustering, i.e.

- one chosen vertex out of V for the case of one-way clustering (L1), i.e. U=(V1) where V1 belongs to V
- two chosen vertices out of V for the case of co-clustering (L2), i.e. U=(V1,V2) where V1 and V2 belong to V but V1 not equal to V2

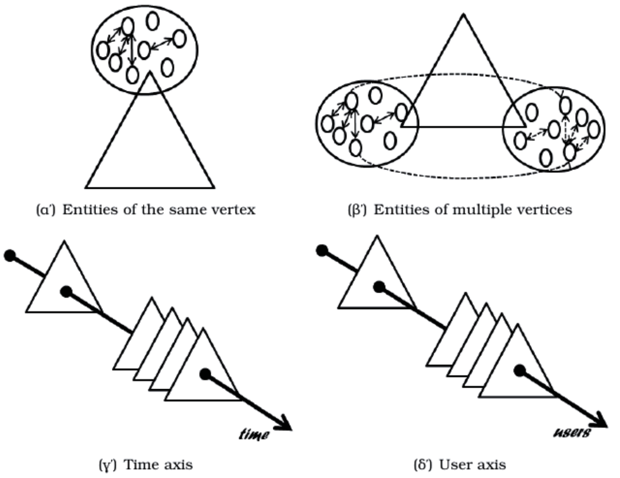

Figure 5: First-level within-vertex relationships (α), Second-level within-two-vertex relationships (β), extension across temporal and user axis (γ and δ).

Step 3) Form **similarity spaces within each vertex** of the semiotic triangle: Introduce symmetric similarity/distance metrics, one for each vertex of the semiotic triangle, if needed, i.e. the tag distance dT(T1,T2), the concept distance dC(C1,C2), and the resource distance dR(R1,R2). Note that here by "symmetric" we are not referring to the property f(a,b)=f(b,a); instead, we are referring to the fact that the distance d accepts two arguments of the same type, for example d(tag1,tag2) is symmetric while d(tag1,object1) is not.

Step 4) Introduce the **six transformation mappings** across the three vertices of the semiotic triangle, i.e. T(R), T(C), R(T), R(C), C(T), C(R), where for example C(T) refers the set of concepts C that is related to a specific tag or set of tags T

Step 5) Introduce **generalized distances**, depending on whether the case is L1 or L2:

a) For the case of one-way clustering L1 using the vertex V1, the generalized distance used for clustering is:

$$dG(V1,V1) = w1*dV1(V1,V1) + w2*dV2(V2,V2) + w3*dV3(V3,V3) \quad (1)$$

where V2 = V2(V1), i.e. the set of entities belonging to vertex V2 that arise out of the transformation mapping V2(V1). For example, if V2=C (concepts) and V1=T (tags), then V2(V1) = C(T) = the concepts that correspond to tag T

Then, proceed by clustering according to the distance dG(V1,V1).

b) For the case of two-way clustering L2 using the vertices V1 and V2, the distance needed for clustering is d(V1,V2), which is not a symmetric distance yet, because V1 is not of the same type as V2. Thus, in this case, transform either V1 into V2, or V2 into V1, by using the appropriate transformation mapping (from Step4). Thus:

$$d(V1,V2) = d(V1,V1(V2)) \text{ or } d(V2(V1),V2) \quad (2)$$

For example, if V1=T (tags) and V2=R (resources), then:

$$d(T,R) = d(T, T(R)) \text{ or } d(R, R(T)) \quad (3)$$

which basically means that in order to calculate the distance between a tag T and a resource R, we first: either transform the resource R to the corresponding tags T(R) and then we calculate the inter-tag distance between the tag T and the tags T(R), or we transform the tag T to its corresponding resources R(T) and then we calculate the inter-resource distance of the resource R with the resources R(T). In both options, the important thing is that we calculate an asymmetric distance (i.e. across types that belong to two different vertices of the semiotic triangle) into a symmetric one (i.e. one that accepts two entities of the same type, i.e. belonging to the same vertex of the semiotic triangle).

Now, after having equated d(V1,V2) = d(V1,V1(V2)), we can proceed by calculating the generalized distance between entities of type V1 as we did in case 5a, i.e.

$$dG(V1,V1) = w1*dV1(V1,V1) + w2*dV2(V2,V2) + w3*dV3(V3,V3) \quad (4)$$

After the above steps, one can proceed with one-way or co-clustering the data under examination.

4.2 Application to previous work

We begin illustrating our framework with approaches that perform one-way clustering (L1). As stated in Equation 1, these algorithms may rely on one-level distances or utilize distances in other vertices as well, after transformation, and combine them all via a weighted sum. An approach that uses one-way clustering, relying solely on one-level distances, is presented in [17], where the authors attempt to extract strongly-related tags based on their co-occurrence. This is the simplest case generated by our framework: V=T (only tags are used), so and the tag distance d(T,T) uses: (w1, w2, w3) = (1, 0, 0). Similar papers include [18] and [19], which again use V=T with one-level distances, but differ from [17] as [18] also uses a second-stage of clustering across users, and [19] uses a different method for co-occurrence calculation and attempts to derive an ontology from the results. Also, there exist some approaches that uses a V different than T. For example, the tool presented in [20] rely on image analysis algorithms for mining and apply them on Flickr images. In this case, V = R.

At the next level of complexity, there exist papers using one-way clustering but utilizing second-level distances that involve two vertices of the semiotic triangle. For example, an approach that uses one-way clustering, but exploits also two-level distances is presented in [21] and [22]. In these particular cases, the authors utilize jointly tag co-occurrence and visual features (i.e. resource characteristics) to estimate tag distance and extract tag clusters. The general idea is that sometimes relying solely on one vertex, such as estimating solely tag co-occurrence, may ignore relationships between entities that do not co-occur. In these cases supplementary knowledge from another vertex, such as visual features of the resources or semantic links through Wordnet or DbPedia, may improve the clustering process. More examples include [23] which uses tags and concepts (through wordnet), and [24] which utilizes tags and Dbpedia.

Moving to more complex approaches, we enter cases which use co-clustering (L2). For example, in [25], an algorithm utilizing both tags (V1=T) and resources (V2=R) is described. In order to perform co-clustering, a non-symmetrical distance function needs to be formed, having heterogeneous arguments (distance between a Tag T and a Resource R). In this case:

d(T,R) = d(T, T(R)) which is more specifically implemented as:

$$\begin{aligned} d(T, T(R)) &= d(T, T1\ldots Tn) = \\ &= \max(dG(T,T1), dG(T,T2), \ldots dG(T,Tn)) \end{aligned} \tag{5}$$

Thus, the transformation mapping T(R) is utilized in order to replace a resource R with the set of tags T1…Tn that corresponds to it, and then the distance chosen is implemented as a maximum operator between the distances of the tag T with each of T1…Tn that correspond to the resource R. Thus, a tag-to-resource distance is calculated in terms of tag-to-tag distances. Those are in turn implemented using a generalized second-level distance utilizing both tags T and concepts C in the following way:

$$dG(T,T) = w1*d(T,T) + w2*d(C(T),C(T)) \tag{6}$$

I.e. a second-level distance using a weighted sum of tag-to-tag distance with the corresponding concept-to-concept distance is utlized, using wordnet for the calculation of conceptual distances. Yet another case L2 (co-clustering) is [26], where two vertices are used for the co-clustering: tags and resources (artists). Numerous other such examples exist.

Finally, moving beyond single-vertex and dual-vertex one way clustering, and also beyond co-clustering, there exist methods that extend the semiosis across users and across the temporal axis. For example, although in [18] the first stage of clustering uses tags-only, at the second stage the user axis is utilized. In numerous other papers the temporal axis is also taken into account for clustering, for example in [28].

5. DISCUSSION

In the previous subsection, the framework presented was used in order to systematically analyze and classify existing work towards social media clustering, under a unifying viewpoint. Examples ranging from single-vertex (tag-only) one-way clustering to multi-vertex co-clustering with second-level distances, where all three of (T,R,C) are utilized, where shown to be special cases of the generic framework. Furthermore, cases involving successions of multiple semiotic triangles across the user- and temporal-axis were covered.

Using the above framework though, one can not only cover and classify existing work, but can also produce novel combinations that fall within the generative power of the framework. For example, one could create novel methods by choosing appropriate similarity metrics within each vertex, choosing subsets of vertices in order to create generalized weighted distances that contain similarities arising across more than one vertex (for example, one could use the triple combination tags – concepts – auditory features of resources), and perform either one-way clustering, or co-clustering, or even extend to higher-dimensional tensor-based methods, such as [27]

There are numerous ways in which our framework can be further refined and extended. For example, one could try to provide more detail treatment of the user- and temporal-axis, or could also try to incorporate other extension axis, Also, moving beyond clustering, other data mining methods such as classification or regression could be covered, so that the framework is further generalized, and thus obtains wider coverage.

6. CONCLUSION

In this paper, we have presented work towards a framework for social semiotic mining. We started by the following motivation: Despite the growth of social media and the direct relevance of semiotics, which is a theory with a long history, no framework has so far been provided, which not only enables the re-examination of social content and tagging under the light of semiotics, but can also be used to analyze data mining and clustering algorithms utilized on social data. Thus, after introducing relevant concepts and examining background work, we provided the outline of such a framework in this paper. Furthermore, we demonstrated how it can be applied not only in order to analyze specific algorithms, but also in order to structure the general space of potential algorithms for clustering data derived from social media, and discussed various possibilities for future extensions, towards the ultimate goal of deriving knowledge from social media and utilizing them in multiple ways beneficial to the common good.

7. ACKNOWLEDGMENTS

The research leading to these results has received partial funding from the Greek General Secretariat for Research and Technology and from the European Regional Development Fund of the European Commission under the Operational Programme “Competitiveness and Entrepreneurship” (OPCE II) - Priority Axis 1 “Promotion of innovation, supported by research and technological development” and under the Regional Operational Programme Attica - Priority Axis 3 “Improving competitiveness, innovation and digital convergence”.

8. REFERENCES

[1] Surowiecki, J. 2005. The wisdom of crowds. Random House LLC.

[2] Al Falahi, K., Mavridis, N., & Atif, Y. 2012. Social Networks and Recommender Systems: A World of Current and Future Synergies. In Computational Social Networks (pp. 445-465). Springer London.

[3] Aitamurto, T. Crowdsourcing for Democracy: New Era In Policy–Making. Publications of the Committee for the Future, Parliament of Finland, 2012. ISBN 978-951-53-3459-6

[4] Al Balooshi, N., Mavridis, N., & Al Qirim, N. A survey on Social Networks and Organization Development, in Proceedings of ACM/IEEE International Conference on Collaboration Technologies and Systems, May 2012, pp. 539-545

[5] Peirce, C. S. 1974. Collected papers of Charles Sanders Peirce: : Science and Philosophy. Reviews, Correspondence, and Bibliography v. 1-8. Belknap Press of Harvard University Press, 1958. (See p. 28).

[6] Ogden, C. K., & Richards, I. A. 1923. The meaning ofmeaning. A Study of the Influence of Language upon

Thought and of the Science of Symbolism. 10th ed. London: Routledge & Kegan Paul Ltd., 1923. (See p. 28).
[7] Chandler, D. 2007) Semiotics: the basics. Routledge
[8] Aristotle. 2011. Categories.On Interpretation. Prior Analytics, Loeb Classical Library No. 325, Harvard University Press
[9] Daoutis, M. 2012. Knowledge Based Perceptual Anchoring: Grounding percepts to concepts in cognitive robots, PhD Thesis, Oerebro Studies in Technology 55, Oerebro 2012.
[10] Ullmann, S. 1972. Semantics: An Introduction to the Science of Meaning. Blackwell paperbacks.
[11] Harnad, S. 1990. The symbol grounding problem. Physica D: Nonlinear Phenomena 42, 335-346, (See pp. 2 sq., 26 sqq., 34, 81).
[12] Vogt, P. 2002. The physical symbol grounding problem. Cognitive Systems Research 3, pp. 429-457.
[13] Van Leeuwen, T. 2004. Introducing social semiotics: An introductory textbook. Routledge.
[14] Carter, B., & Knight, D. (2008). Semiotic domains and nontextual technologies. Retrieved at 2013-11-19, from http://etec.ctlt.ubc.ca/510wiki/Semiotic_Domains_and_Non-Textual_Technologies
[15] Fellbaum, C. 1998. Wordnet: An electronic lexical database. Bradford Books.
[16] Plato. 1989. Timaeus Critias Cleitophon Menexenus Epistles: v. 9 Loeb Classical Library.
[17] Begelman, G., Keller, P., & Smadja, F. 2006. Automated tag clustering: Improving search and exploration in the tag space. In Collaborative Web Tagging Workshop at WWW2006, Edinburgh, Scotland 15-33.
[18] Peter Mika. 2007. Ontologies are us: A unified model of social networks and semantics. Web Semant. 5, 1 (March 2007), 5-15. DOI=10.1016/j.websem.2006.11.002 http://dx.doi.org/10.1016/j.websem.2006.11.002
[19] Schmitz, P. (2006, May). Inducing ontology from flickr tags. In Collaborative Web Tagging Workshop at WWW2006, Edinburgh, Scotland (Vol. 50).
[20] Bumgardner J. 2006. Experimental colr pickr. Available at: http://www.krazydad.com/
[21] Giannakidou, E., Kaklidou, F., Chatzilari, E., Kompatsiaris, I., & Vakali, A. (2010). Harvesting Intelligence in Multimedia Social Tagging Systems. In Emergent Web Intelligence: Advanced Information Retrieval (pp. 135-167). Springer London.
[22] Aurnhammer, M., Hanappe, P., Steels, L., Augmenting navigation for collaborative tagging with emergent semantics. In Proc. of the 5th ISWC (2006)
[23] Börkur Sigurbjörnsson and Roelof van Zwol. 2008. Flickr tag recommendation based on collective knowledge. In Proceedings of the 17th international conference on World Wide Web (WWW '08). ACM, New York, NY, USA, 327-336. DOI=10.1145/1367497.1367542 http://doi.acm.org/10.1145/1367497.1367542
[24] Anastasia Stampouli, Eirini Giannakidou, and Athena Vakali. 2010. Tag disambiguation through Flickr and Wikipedia. In Proceedings of the 15th international conference on Database systems for advanced applications (DASFAA'10), Masatoshi Yoshikawa, Xiaofeng Meng, Takayuki Yumoto, Qiang Ma, Lifeng Sun, and Chiemi Watanabe (Eds.). Springer-Verlag, Berlin, Heidelberg, 252-263.
[25] Giannakidou, E., Koutsonikola, V., Vakali, A., & Kompatsiaris, Y. 2008. Co-clustering tags and social data sources. In Web-age information management, 2008, WAIM'08. The ninth international conference on. 317–324.
[26] Li, J., Li, T., & Ogihara, M. (2010). Hierarchical co-clustering of music artists and tags. In 11th International Society for Music Information Retrieval Conference. Utrecht, Netherlands:[sn] (pp. 249-254).
[27] Nanopoulos, A., Gabriel, H., & Spiliopoulou, M. (2009). Spectral clustering in social-tagging systems. In 10th int. conf. on Web information systems engineering (pp. 87–100).
[28] Becker, H., Naaman, M., Gravano, L. (2010). Learning similarity metrics for event identification in social media. In WSDM '10: Proceedings of the third ACM international conference on Web search and data mining (pp. 291–300). New York: ACM.