On Building Semantically Enhanced Location-Based Social Networks

Cláudio de Souza Baptista, Luciana Cavalcante de Menezes, Maxwell Guimarães de Oliveira, Ana Gabrielle Ramos Falcão and Leandro Balby Marinho

Information Systems Laboratory – Federal University of Campina Grande (UFCG)
Av. Aprígio Veloso 882, Bloco CN, Universitário – 58.429-140
Campina Grande – PB – Brazil
{baptista, lbmarinho}@dsc.ufcg.edu.br, maxwell@ufcg.edu.br,
{lucianacm, anagabrielle}@copin.ufcg.edu.br

Abstract. Currently, most social content sites enable users to enrich their tags with semantic metadata, such as geographic metadata in the case of Location-Based Social Networks (LBSN). However geographic metadata alone only unveils a very specific facet of a tag, leading to the need for general purpose semantic metadata. This paper introduces DYSCS – Do it Yourself Social Content Sites – a platform that combines Web 2.0 and Semantic Web technologies for assisting users in creating their own LBSN enriched with semantics. DYSCS improves information organization and retrieval and provides distinct functionalities such as multimodal interface. Because of its ontology driven architecture, DYSCS is highly reusable and interoperable.

1. Introduction

Social content sites are Web 2.0 applications that empower ordinary users with the ability to create, edit, publish, and annotate online Web content [John 2013]. Currently, there are many social content sites, each one driven to a specific interest or domain. For example, users can contribute and share videos in YouTube, photos in Flickr, news in Digg, and musical preferences in Last.fm. With the growing popularity of these sites, the amount of shared information has increased significantly, which, in spite of leading the Web into a more open and democratic environment, put forward new challenges for organizing and retrieving information.

A popular way of organizing shared information in social content sites is to leverage the “wisdom of the crowd” approach. In this approach, typically known as collaborative tagging, users can assign a set of freely chosen keywords, called tags, to any given resource (or content). These tags act as indexes for resources and are used for information retrieval.

Many social content sites enable users to take into consideration the geographic context of resources through geographic annotations (or geotagging) [Naaman 2011]. These social content sites enhanced with geographic information raise a new concept called Location-Based Social Networks (LBSN). LBSN add a new dimension to social networks by introducing Geo-tagged user-generated media, such as texts, photos, and videos. Hence, users annotate their resources with geotags, which are tags associated with geographic metadata, e.g., the latitude and longitude coordinates of the resource
being annotated. LBSN are becoming increasingly common nowadays, mainly due to cameras and smartphones that come equipped with embedded GPS devices, generating high volume of georeferenced multimedia data. Moreover, many LBSN already provide tools to facilitate enhancing tags with geographic metadata, for example, through Web interfaces enriched with interactive maps. Geographic annotations enable users to better define the geographic scope of their resources, thereby improving information retrieval regarding specific locations.

A well-known drawback of using tags for organizing and retrieving information is the lack of well-defined semantics, which can lead to problems such as: polysemy, synonymy or misspellings [Chen et al. 2012]. These problems might reduce the user's ability to find information. For example, a user who is searching for the Java programming language may get resources about the Java Island instead, since both are annotated with the tag Java (polysemy). In order to mitigate these problems, it is necessary to ensure an annotation that enables to enrich tags with well-defined semantics. To this end, the social semantic annotations arise, which are user-generated tags enriched with machine processable semantic metadata. Enriching social annotations with semantic metadata allows humans and machines to process the meaning of tags efficiently, hence improving, eventually, their ability to find relevant information.

Most of the existing social content sites available employ social tagging of resources through either regular social annotations (i.e., non-semantic) or social geographic annotations. However, social semantic annotation is still rarely featured in these sites. Moreover, we observed that geographic metadata is underexploited in the LBSN, such as Flickr, for example, in which the geographic search is restricted to names of places and visualization of shared resources on a map. To better exploit the geographic metadata, spatial topological functions, such as buffer, contains and not contains, could be used in order to define, precisely, the geographic search scope, allowing, thus, the retrieval of more precise results. For example, a query about bookstores within a radius of 20 km from a given point, could be fulfilled through the function buffer, while a query about gas stations in a given region marked on the map, could be fulfilled through the spatial function contains.

In order to address the aforementioned problems, we introduce the DYSCS - Do it Yourself Social Content Sites – platform, that combines technologies from Web 2.0, Semantic Web and advanced exploitation of geographic metadata, for assisting users in creating LBSN. The platform uses an ontology-based approach to model LBSN and user interactions. In order to organize and retrieve information on the LBSN built with the platform, a strategy that combines social semantic and geographic annotations is used. Thus, tags can be associated with both general purpose semantics and geographic metadata.

The DYSCS platform was developed with the aim of assisting users in creating LBSN, focusing on improving the organization and retrieval of information through the use of semantic and geographic metadata. The platform is based on an underlying ontology that models the interactions between users, multimedia resources, semantic and geographic tags, and other features of the site, such as description and geographic scope. Ontologies have appealing features that can improve the overall quality of the
application such as interoperability, reusability and they are highly processable by machines [Simperl 2009].

Social content sites may be used as a communication channel between the society and government agencies. An example with this purpose may be cited, where an individual can use the DYSCS platform to create a site so that people can report street problems they find on the streets of the city of Rio de Janeiro, Brazil. The citizens, then, are able to mark on the map the places where they witnessed problems such as, graffiti, bad lighting or trash on the streets and also share images regarding the problem. An advantage of a social content site created with DYSCS is the ability to use geographical and semantic tags to organize the publications in the site and improve the information search.

The main contributions of DYSCS are: ontology built for LBSN, definition of geographic scopes of a LBSN, addition of semantic tags, addition of GeoTags, and a web-based multimodal interface.

The remainder of this paper is structured as follows. Section 2 presents the OntoDYSCS ontology. Section 3 focuses on DYSCS architecture. Section 4 addresses the use of the DYSCS platform. Section 5 discusses related work. Finally, section 6 concludes the paper and highlights further work to be undertaken.

2. ONTODYSCS: Ontology for the DYSCS Platform

This section presents OntoDYSCS, an ontology developed to serve as a model for the DYSCS platform. The main goal of this ontology is combining social semantic and geographic annotations.

A base ontology is used for creating other ontologies, as it contains concepts and instances that can serve multiple domains. According to Simperl (2009), reusing ontologies can reduce the development costs since it avoids re-implementing components already available, which could be incorporated into another ontology, after eventual minor adjustments. Thus, OntoDYSCS extended and reused several base ontologies for the representation of people, online communities, geographical elements, date, time and digital media.

To store the information of the DYSCS platform (such as the social content sites and the semantic and geographical tags) and to perform the inference on such data, the Jena Framework is used. When using Jena, a Semantic Web language needs to be chosen and a Model object created, responsible for storing a reference to the Semantic Web data. RDF, RDFS and OWL are examples of possible Semantic Web languages, and the latter was chosen for our platform.

For representing the semantic aspects of DYSCS LBSN, the MOAT (“Meaning of a Tag”) ontology, proposed by Passant and Laublet (2008), was used. MOAT is an extension of the “Tag Ontology”, which in turn, is based on the SIOC (“Semantically Interlinked Online Communities”) [Breslin et al. 2005] and FOAF (“Friend of a Friend”) ontologies. The MOAT ontology enables meaning to be added to the regular user generated tags through the moat:tagMeaning property (from the Meaning class), hence creating the semantic tag.
For modeling the geographic aspects of a LBSN, we used GeoOWL, a W3C ontology that uses the GeoRSS vocabulary for modeling classes and geographic properties. This ontology supports the concept of points (i.e., a pair of coordinates), lines (i.e., one or more pairs of coordinates), boxes (i.e., exactly two pairs of coordinates), and polygons (i.e., at least four pairs of coordinates).

The class `tags:RestrictedTagging` is used to represent a social annotation performed by a specific user to a specific resource, assigning to such resource a tag, a semantic metadata and/or a geographic metadata.

The class `tags:RestrictedTagging` is originated from the "Tag Ontology" and its main properties are:

- `tags:associatedTag`: associates a tag to the social annotation. The tag is modeled by the class `moat:Tag`;
- `tags:taggedResource`: identifies a resource that is being annotated. The class used to model the resource must be a subclass of `owl:Thing`;
- `moat:tagMeaning`: assigns a meaning to the tag used in the annotation. This meaning is represented by the class `scs:MeaningURI`;
- `tags:taggedBy`: identifies the person creating the annotation. The class that represents the person must be a subclass of `foaf:Person`; and
- `scs:has_location`: associates in the annotation a geographical location, for such, `gml:_Geometry` objects are created.

In OntoDYSCS, the semantic tag is derived from the social annotation. Hence, a tag is called a semantic tag if the properties `tags:associatedTags` and `tags:tagMeaning`, from the social annotation it belongs to, are filled. In this case, the ontology provides a semantic social annotation (Figure 1 (a)). The same way, the geographic tag must have in its social annotation the properties `tags:associatedTag` and `scs:has_location` - if this happens, the ontology provides a geographic social annotation (Figure 1 (b)). In the case the tag has all three properties, it is considered a semantic and geographic tag and its social annotation is called a semantic and geographic social annotation (Figure 1 (c)).

In order to allow the modeling of semantic and geographic annotation, we extend the Passant and Laublet model (MOAT ontology), which is based on quadruples of the form (user, resource, tag, meaning of tags), with geographic metadata (GeoOWL ontology).
Besides these ontologies, we used FOAF to represent users and their social networks, and SIOC to integrate the published information into the LBSN site.

3. DYSCS: Do it Yourself Social Content Sites Platform

This section presents DYSCS, a platform proposal for assisting users in creating LBSN that uses OntoDYSCS ontology and combines technologies from Web 2.0, Semantic Web and advanced exploitation of geographic metadata.

In DYSCS, when a user annotates a resource with a regular tag, he can further extend this annotation with semantic and geographic metadata. He can also annotate a resource using only semantic or geographic tags.

In order to create a semantic tag, the user must choose a “meaning” for the tag. For choosing tag meanings, DYSCS provides an approach based on the principles of Linked Data, in which URIs of existing resources are used to define the tag meanings. These URIs are retrieved from the Freebase database.

When the user chooses a topic from Freebase data repository, the system associates the URI related to that topic to the tag chosen by the user. Next, this semantic annotation is stored in the knowledge base of the DYSCS platform, which contains the relationships between the user, resource, the tag, and its meanings.

The assignment of geographic metadata (latitude and longitude coordinates), in order to generate the geotags, is facilitated by the mashup of Google Maps and also by the latitude and longitude properties of Freebase topics of the type location.

Figure 2 depicts the DYSCS interface that allows the assignment of semantic and geographic metadata. The text fields ‘Location’ and ‘Semantic’ use an auto-complete feature that displays the topics extracted from the Freebase database referring to the word typed in.

![Figure 2. Interface for creating geotags, semantic tags, and the combination of both.](image)

When the user types a word in the field ‘Location’, Freebase topics related to the type location are displayed. After selecting the topic, the semantic and geographic tag manager is triggered in order to retrieve the latitude and longitude coordinates of the
topic. If they are found, the point is marked on the map so the user can visualize the results.

When the coordinates of the place are not found in Freebase, the system uses the Google Geocoding API for trying to retrieve this information. However, before using the API, the names of the city and the country containing the place of interest is added to the name of place of interest. This basic information about the place of interest is retrieved from Freebase and is used aiming to increase the chance for the algorithm from Google to retrieve the coordinates of the place correctly.

For example, suppose that Freebase does not have the coordinates for ‘Sugarloaf Mountain’, then the names of the city and the country containing the ‘Sugarloaf Mountain’ is added, resulting in ‘Sugarloaf, Rio de Janeiro, Brazil’. Next, using the Geocoding API, we try to retrieve the latitude and longitude for this place. Finally, if none of the aforementioned strategies work, the user can mark the coordinates of the place of interest in the map.

The algorithm developed for creating and maintaining social semantic annotations in the DYSCS platform is shown in Figure 3.

```
createDYSCsSocialAnnotation(tag, semantic, location, resource, user) {
    RestrictedTagging socialAnnotation = OntoDYSCSManipulator.createSocialAnnotation;
    socialAnnotation.addTag = tag;
    socialAnnotation.addThing = resource;
    socialAnnotation.addPerson = user;
    if (semantic != null) then
        MeaningURI semantic = OntoDYSCSManipulator.createSemantic;
        socialAnnotation.addSemantic = semantic;
    if (location != null) then
        _Geometry location = OntoDYSCSManipulator.createLocation;
        socialAnnotation.addLocation = location;
    PlatformManipulator.addDYSCsSocialAnnotation = socialAnnotation;
}
ontoDYSCsSearch(what, wherer, who, when, tag) {
    OntoDYSCSManipulator = OntoDYSCSManipulator.instantiateModel;
    Query query = OntoDYSCSManipulator.createQuery(what, wherer, who, when, tag);
    Response response = OntoDYSCSManipulator.executeQuery(query, model);
    while (response.hasNext)
        // process the result to the client
}
```

Figure 3. The DYSCS Algorithm.

3.1. Architecture

The Model-View-Controller (MVC) design pattern was adopted in the development of the DYSCS architecture. Such architecture is composed of three layers: visualization, control and persistence. Figure 4 depicts the three-layer based architecture of DYSCS.

The visualization layer deals with the elements of the application that are visible to the end users, i.e., the interface between the system and the users. In DYSCS, users can access the visualization layer for either building LBSN sites or contributing with information on existing ones.
The control layer is responsible for processing the operations required by users, and contains the logic of the system. As Figure 4 shows, the DYSCS control layer is composed of seven main modules: the platform manager, the LBSN manager, the user’s manager, the multimedia resources manager, the semantic and geographic tag manager, the information persistence module, and the information discovery module.

The platform manager module represents the kernel of the DYSCS platform and is responsible for managing the social content sites built with the platform, and also users who register with the aim of owning a social content site. This module receives the necessary information from the visualization layer in order to create, remove, edit, or search for a site or site owner, verify the consistency of the information and pass it to the module responsible for realizing the desired action.

The LBSN manager module is responsible for managing the content of all the LBSN sites created through DYSCS. Moreover, it manages the searches for information contributed by users.

The user’s manager module manages the site owners, authenticated and anonymous, and their social content sites. This module interacts with the multimedia resources manager module, in order to store the photos that identify the users; with the information persistence module, in order to persist the information about users; and with the information discovery module, used for creating, editing, and authenticating users.

The multimedia resources manager module manages the multimedia resources of the social content sites. This module interacts with the information persistence module in order to save the existing multimedia files in a multimedia repository. The multimedia resources manager also communicates with the information discovery module, in order to conduct searches for multimedia resources.

The semantic and geographic tag manager module deals with the semantic and geographic tags in the DYSCS platform. The information persistence module, in turn, is responsible for providing data persistence services. This module has communication interfaces with the DYSCS persistence layer. Finally, the searches realized in the LBSN are processed by the information discovery module, which also communicates with the persistence layer through several interfaces.
Finally, the persistence layer contains the data that will be processed and/or generated by the control layer and visualized by the user in the visualization layer. We use three data repositories in the DYSCS platform. The first one is used to store the information related to the LBSN sites, according to the OntoDYSCS ontology. The control layer communicates to this repository through the Jena Framework. The second is a file repository, used to store multimedia files shared by the users, and pages containing the templates of each existing LBSN. The third and last repository is a database created to support the spatial functions and is used to store the geographic metadata of geotags and the geographic scope of the LBSN sites. In this repository the spatial functions required by the users through the visualization layer are processed.

4. A DYSCS LBSN Instance

The creation of a LBSN site through DYSCS involves user authentication, choosing a name for the site, configuring general information about the site, defining the geographic scope, and creating map markers. The entire creation process is executed through Web interfaces and the utilization of the underlying ontology is transparent to the user. Finishing these steps, the site is ready for use.

Figure 5 presents the final interface of the LBSN site built with the purpose of sharing information about the urban problems of the city of Rio de Janeiro, Brazil. Note that there is a polygon in the map (Figure 5 (a)), which is used to delimit the geographic scope chosen for the site. Therefore, any information that contains geographic metadata must be within the geographic scope, otherwise it cannot be published into the site. The geographic scope can only be defined by the owner of the site, who can edit it anytime, however, care must be taken since changing the geographic scope can lead to loss of shared information that might be out of the limits of the new geographic scope.

For conducting tag-based search, the user must use the text field shown in Figure 5 (b). For advanced search, the user must access the respective Web page by clicking in the link ‘Advanced Search’, positioned at the right hand side of the tag-based search text field (Figure 5 (b)). The advanced search allows the user to employ the spatial operators contains, not contains, and buffer, in order to better filter the results, and also allows to filter resources according to the people who published it. Furthermore, it is possible to combine semantic, geographic, and social elements in the search. The results of performed searches can be visualized in the interactive map or in a new window with textual information.

For sharing information, both authenticated and anonymous users are allowed. Figure 5 (c) shows a list of map markers that can be used for sharing information directly on the map. These markers are defined by the site owner. In the specific case of the social content site depicted in Figure 5, the map markers, among others, are: trash, public lighting, hole, graffiti, and sewage. Figure 5 (d) depicts the sharing area and the searching for information published in the site that does not use map markers. This information can be annotated with semantic and/or geographic tags during publishing.

The members of the LBSN community can be visualized in the area shown in Figure 5 (e). Moreover, registered users can search for other users and establish friendship relations with them.
Figure 5. LBSN site created with the DYSCS platform.

For conducting advanced searches in sites created with DYSCS, users should access the advanced search interface. In this interface, users can configure several parameters in order to obtain the desired information. An user of the site ‘Street problems in Rio de Janeiro’ can search, for example, for the posts annotated with the semantic metadata ‘Asphalt concrete’ and with geographic metadata within a buffer of 2.5 km radius from the location named ‘Copacabana Beach’. The results of this search are displayed in textual format and the ones among those associated with geographic metadata can be visualized in the map.

Several types of queries can be submitted. In general, users can personalize their queries by:

- **Regular Tags**: The user informs the tags that will be used by the search engine. The system returns as results all the resources that were annotated with these tags;

- **Semantic Tags**: In this option, the user must inform the tags and their associated meanings to the search interface. The system then returns all the resources annotated with these tags that are associated with the given meanings. The user can also conduct searches passing to the system only the meanings of the resources, without explicitly specifying the tags associated to these meanings. In this particular case, all the resources annotated with the tags associated with the informed meanings are returned;

- **Geographic Tags**: The user informs the tags and their geographic coordinates or the name of the place to the search interface. He can use spatial operators, such as buffer, contains, and not contains to refine the search. The result is shown in a
mashup of Google Maps. As well as with semantic tags, it is possible to conduct searches using only geographic metadata, without having to explicitly inform an associated tag;

- **Semantic and Geographic Tags**: The user can search for resources informing both semantic and geographic tags to the search interface. To this end, the user is required to fill up the fields corresponding to the tag and its semantic and geographic metadata;

- **Members and Date**: The user can search for resources based on the users (members or anonymous) who publish it, the date when the resource was shared, and also on the type of information published.

5. Related Work

Currently, there are several applications built with Web 2.0 technologies such as blogs, wikis, social network sites, social content sites and also software platforms that assist users in creating their own LBSN. Common to all these applications is the fact that the information available in the system is contributed by the users. The utilization of social annotations to promote information sharing among users is a common practice in the context of the Web 2.0 paradigm.

Marchetti et al. (2007) introduce SemKey, a semi-automatic approach to enrich tags with semantic metadata. In their work, users are responsible for assigning semantics to a tag. DYSCS uses the same principle, where users can enrich tags with semantic metadata. However, additionally, DYSCS's users can also assign geographic metadata to their tags.

Regarding the development of social networks, Web 2.0 platforms such as Ning and Elgg provide customized services for the users to create their own networks, with functionalities like multimedia sharing and discussion forums. Nevertheless, they are limited in the sense that neither of them support semantic or geographical tags. Additionally, DYSCS distinctly provides several other useful functionalities such as the definition of a geographical scope for the site, the ability to perform semantic or spatial searches on the available resources and multimodal interface.

Mahmud et al. (2010) focus on using social network and user’s context (e.g. user location), in a ‘Help-me!’ scenario in a vehicular network. They obtain timely and relevant information from close peers, considering their areas of expertise and spatial closeness.

Still regarding user's context, another relevant project is the WeGov [Wandhöfer et al. 2012], which addresses the citizens connection with governmental policy makers through popular social networks, such as Twitter and Facebook, showing the retrieved information on a map. The geographical information, though, is also not fully explored.

Bilandzic et al. (2008) present a system enabling visitors and new residents in a city to obtain the knowledge and experiences of local residents. That system aims to facilitate social navigation in urban places. Lee and Sumiya (2010) developed a geo-social local event detection system by monitoring crowd behaviors from geo-tagged microblogging sites, such as Twitter, using posts about such events. Doytsher et al.
present an integrated socio-spatial graph to monitor life patterns through the integration of social network and spatial network graphs.

Baykurt (2011) discusses the FixMyStreet LBSN, where UK citizens can report problems about their streets - such as bad pavement or lack of lighting - and this interaction is done through a map.

Santos and Furtado (2012) proposed a service-oriented architecture called SeMaps that makes it possible to generate semantic crowd maps "that have the power to perform inferences and/or access external sources that constitute useful and appropriate information to the map context". This approach also explores the semantic value of the information, other than the geographic metadata, and was integrated with Wikimaps in order to see it in action.

6. Conclusion and Future Work
Currently, there is a strong trend towards LBSN, in which users can contribute and share georeferenced content. Most of these sites still have lack of semantics of user tags, which can lead to poor annotations and retrieval of information. Furthermore, it is important to provide more sophisticated geographic search tools, which can reduce precision.

In this paper we introduced the DYSCS platform, which aims at addressing the aforementioned limitations of existing LBSN sites. DYSCS aims to help users in the creation of geo-social content sites enhanced with general purpose semantic metadata, improving the organization of the sites information and the complexity of searches.

In order to facilitate the creation of semantic metadata, we developed an interface that access the Freebase repository for retrieving URIs that define the existing resources. The creation of geographic tags was also facilitated by Freebase, through the retrieval of the latitude and longitude coordinates from topics of the type location, and the mashup of GoogleMaps, that enables users to interact with a map to define the metadata of the geotag.

As future work, we intend to incorporate trust features, concerning the shared information, into OntoDYSCS; to improve the search for textual documents through traditional keyword searching using inverted indexes of terms; and to add recommender system services in order to suggest users, resources, and tags that match user preferences. Moreover, we intend to conduct usability tests to assess the user satisfaction regarding the Web interface of the LBSN sites built with DYSCS.

7. Acknowledgments
The authors are grateful to the Brazilian Research Council - CNPQ - for funding this research.

References


