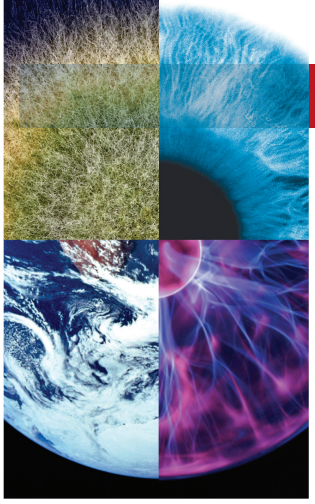

Editor: **Jim Hendler,** Rensselaer Polytechnic University, hendler@cs.rpi.edu

Open Government and Citizen Participation in Law Enforcement via Crowd Mapping

Vasco Furtado, Carlos Caminha, Leonardo Ayres, and Henrique Santos, *Universidade de Fortaleza, Brazil*

Crowd mapping, an activity that combines aggregated Geographic Information System-generated maps on the Web with crowd-generated content, flourishes daily.[1,2] Sites such as Wikimapia (www.wikimapia.com), Click2fix (www.click2fix.co.sa), Crowdmap (www.crowdmap.com), and OpenStreetMap (www.openstreetmap.org) empower citizens to create a global patchwork of geographic information, while Google Earth and other virtual globes encourage volunteers to develop interesting applications using their own data. In crowd-map applications, the digital map works as a blackboard that accommodates stories told by people about events they want to share with others, typically via their social networks.

In parallel with this wave of citizen participation is an explosion of access to huge amounts of data on open source and open data networks. This increasingly democratized access to information has led to greater civic education, which in turn produces better-informed citizens and communities.[3] But even though people in countries such as the US and the UK enjoy an unprecedented ability to access, say, crime statistics, such information typically isn't available in other countries. Instead, there persists a culture with a lack of transparency, and, consequently, few services that use open data are available to citizens.

It's in this particular context that we initiated a project called WikiCrimes (www.wikicrimes.org) five years ago in Brazil. It's driven by three goals: to give more transparency to criminal information, to provide means for citizen crime prevention, and to reduce the phenomenon of underreported crime (those that aren't reported to law enforcement). Several countries around the world share these goals, particularly those in which the population suffers high rates of violence.

WikiCrimes

WikiCrimes aims to offer a common interaction space for the general public, where people can report criminal activities as well as keep track of crime locations. We based it on the principle that the citizens holding the most information about a crime are the people affected—victims, witnesses, and so on. If they want to make such information public, they can. What we intend with WikiCrimes is a "global blackboard," where victims, witnesses, or anyone else with information about a crime can report what happened and where to alert others on a scale larger than close social contacts. If WikiCrimes gains active participation, crime mapping could start to happen collaboratively, and soon other citizens would benefit from having access to information about where crimes have occurred.

Project Goals and Motivation

The veracity and accuracy of information about where crimes occur, as well as data on the characterization of such crimes, is typically monopolized by law enforcement agencies and is therefore highly centralized. This monopoly ultimately creates tension between such agencies and the general public, because the former tend to oppose disclosure and transparency. Allied to this context is the general sentiment that law enforcement fails to provide a quality public service (primarily in South America),

1541-1672/12/$31.00 © 2012 IEEE
Published by the IEEE Computer Society

which tends to diminish citizens' trust in those agencies.

These factors have led to the growing problem of underreporting. In Brazil, for instance, it's common to hear that someone was mugged and didn't file a police report because he or she thought it wouldn't matter or bring the criminal to justice. Surveys conducted with crime victims in several Brazilian states show that underreporting in densely populated areas could be as high as 50 percent for certain types of crimes.[4] The consequence is disastrous in terms of public policies and planned police actions because the official crime map reflects a trend that differs dramatically from what actually occurs in real life.

Services Provided to Citizens

To foster participation in WikiCrimes, we designed a plan to establish partnerships with nongovernmental organizations, news corporations, business representatives, municipal governments, and other organized sectors. To do this, we tried different strategies, such as a marketing campaign, which involved distribution of ads via folders, informational pamphlets, and bumper stickers. One good example of a partnership was the one we established with car insurance brokers, who often have information about vehicular crimes and can register them in the system. Unfortunately, open data from police departments in Brazil are rarely available, so most of our initiatives are related to participation by the general public. Through our work, we've defined technological functionalities that can foster further collaboration.

Alert services. Even though the rationale behind WikiCrimes is founded on the notion of solidarity (registering crime to help others), providing apps and services is a crucial strategy for fostering collaboration. In WikiCrimes, we defined a service called WikiCrimes Alert, which a user can subscribe to for alerts via email about crimes reported in a user-defined geographic area. Apps are available for both iPhone and Android devices.

Social networks. Another important strategy for fostering collaboration and advertising the project was to create a means of integrating it into a social network application. To do that, we developed a mini-application following Google's proposal of an Open Social API on Google's Orkut (www.orkut.com), a popular social network application in Brazil, and Ning (www.ning.com), a platform to generate social networks. The idea is to provide a tool for social network users to report crimes and alert their friends about security problems in a particular region. These mini-apps were an opportunity to reach a wider audience and encourage more participation (mainly from youth). A Facebook application is currently being developed.

Online newspapers. Even though Web applications offer a very specific way of advertising, the use of traditional media is fundamental for advertising any public project. We established partnerships with local newspapers, whereby we made it possible for the journalists themselves to generate a widget (an embedding of WikiCrimes) with a small map of crimes that they could insert into any online news display.[5]

Real versus virtual worlds. The lack of signs to alert people about the prevalence of pickpockets is emblematic of Brazil's lack of transparency. Pickpockets are largely encountered in very dense places, such as tourist zones. In WikiCrimes, we gave users the ability to produce their own "beware of pickpockets" sign. They can create a QR code that refers to the crime map describing how dangerous a particular place is.

New types of victim surveys. People who report incidents in WikiCrimes have the option of describing what they perceived as the "causes" of a particular crime, such as poor lighting. Environmental criminologists in particular examine the place and the time when the crime happened. They're interested in land usage, traffic patterns and street design, and the daily activities and movements of victims and offenders.[6] Capturing people's perception about these features and problems associated to them is one way to supply governments with information of areas they could possibly improve.

The Research behind WikiCrimes

More than an innovative technological system, WikiCrimes has proven to be a rich space for research in intelligent systems, ranging from its use of natural language processing (NLP) in online news stories to semantic representations of reported crimes for open government to the creation of methods to identify malicious user activity (for example, someone trying to generate false trends in crowd maps). Here, we review some of the research we're working on related to the challenges we faced in the context of crowd mapping with open data.

Improving Communication between Government and Citizens

For two-way communication and information flows between governments and citizens to be effective, the different data sources must follow a pattern that can, for example, enable reliable comparisons.

For this purpose, we proposed a representation for the concepts of "crime" and "report crime." This representation isn't restricted to open information released by a police department, but some information is mandatory to define a unique instance. For example, a crime will have a type, a date, a time (imported from the time ontology[7]), a precise address (geographical coordination), and a description. Information about the people involved such as perpetrators, witnesses, and victims can also be included but isn't mandatory.

Our crime ontology is basically a hierarchy for inferential purposes and represents various classifications of crime type. We define crime events as specializations of the `Event` class, from the `Event` ontology. According to the `Event` ontology, "an event is an arbitrary classification of a space/time region, by a cognitive agent. An event may have actively participating agents, passive factors, products, and a location in space/time something that can have agents that interacts, produces and location in space and time."[8] To describe where a crime occurred geographically, we use the ontology `wgs84` to express location in terms of latitude and longitude.

Typically, a detailed identification of the people involved isn't open information due to privacy concerns, but rules vary depending on countries and cultures. In Brazil, for instance, the media discloses the names of homicide victims, but in the US, raw crime data doesn't include victims' names.

We defined a crime ontology inspired by the `criminal act` ontology in the context of the OpenCyC project and also took into consideration the US Federal Bureau of Investigation's (FBI's) uniform crime report standard. A crime report by its very nature refers to a particular crime and contains information about the reporting itself, including the reporter's name and the time and date he or she filed the report. Because a crime report contains basic provenance information, we imported the `PML2` ontology,[9] which includes classes and properties to represent source trust and data credibility. These properties are important in the combination of open crime data from a large variety of sources that are sometimes anonymous. The `CrimeReport` class is a subclass of `pmlp:Information`. We also used some specific properties to describe a report such as `pmlp:hasCreationDateTime` (hour of the report), `pmlp:hasDescription` (text of the report), and `pmlp:hasSource` (entity that published the report).

Figure 1 describes the main classes of both ontologies (see www.wikicrimes.org/ontology/ontology.owl for further information[10]).

Determining Credible Information

Crowd-mapping systems face a constant trade-off between diminishing the constraints imposed on users (to increase the number of participants) and imposing rigid control to avoid unwanted behavior (such as the reporting of fake information). Why does that specific area have so many crimes? Is it true, or is someone trying to make a joke, speculate about the real estate market, or diminish the image of the local police?

WikiCrimes doesn't have many prerequisites on its members. The only personal questions asked are name and valid email address; no document identification is required. It's up to the user to provide information that increases his or her crime-reporting credibility in the system, for example, by adding links to videos, newspapers, photos, or any other documents (such as a police report). Moreover, for every criminal fact registered in the system, WikiCrimes requests an indication of at least one person who can confirm that the information posted is true; this is to increase the data's credibility, possibly making the system as a whole more reliable. These confirmations generate a graph where the vertices represent WikiCrimes users and the edges represent other people who can confirm the registered criminal report. By creating this social network, we are able to build a reputation model that plays a strong role in identifying "bad users" in the system.[5] By "reputation," we mean a score that represents the community's view about a fellow member; trust can be calculated upon the acknowledgment of a given reputation.

In WikiCrimes, our social network is made of registered users, forming the Social Network Layer, and the information posted in the system forms the Information Layer. The goal is to build a function that will calculate user reputation and reflect data trustworthiness posted in the Information Layer. Some entities, such as the press and governmental agencies, are labeled as Certifier Entities, and considered to be well reputed. But that isn't enough; the system's openness—the sense that anyone can be a user—doesn't facilitate the task of knowing the reputations of all users.

Reputation attribution to users who aren't qualified as Certifier Entities is fundamental to success. We can assume that a user delegates a commitment to another user when the former indicates the latter to give an opinion about a particular report. The user's acceptance of the commitment indicates a relationship of trust between the agents. If the second user breaks the commitment, he or she will be penalized in reputation points. Certifier Entity users have a good reputation to start with and serve as a starting point for the propagation of trust to the

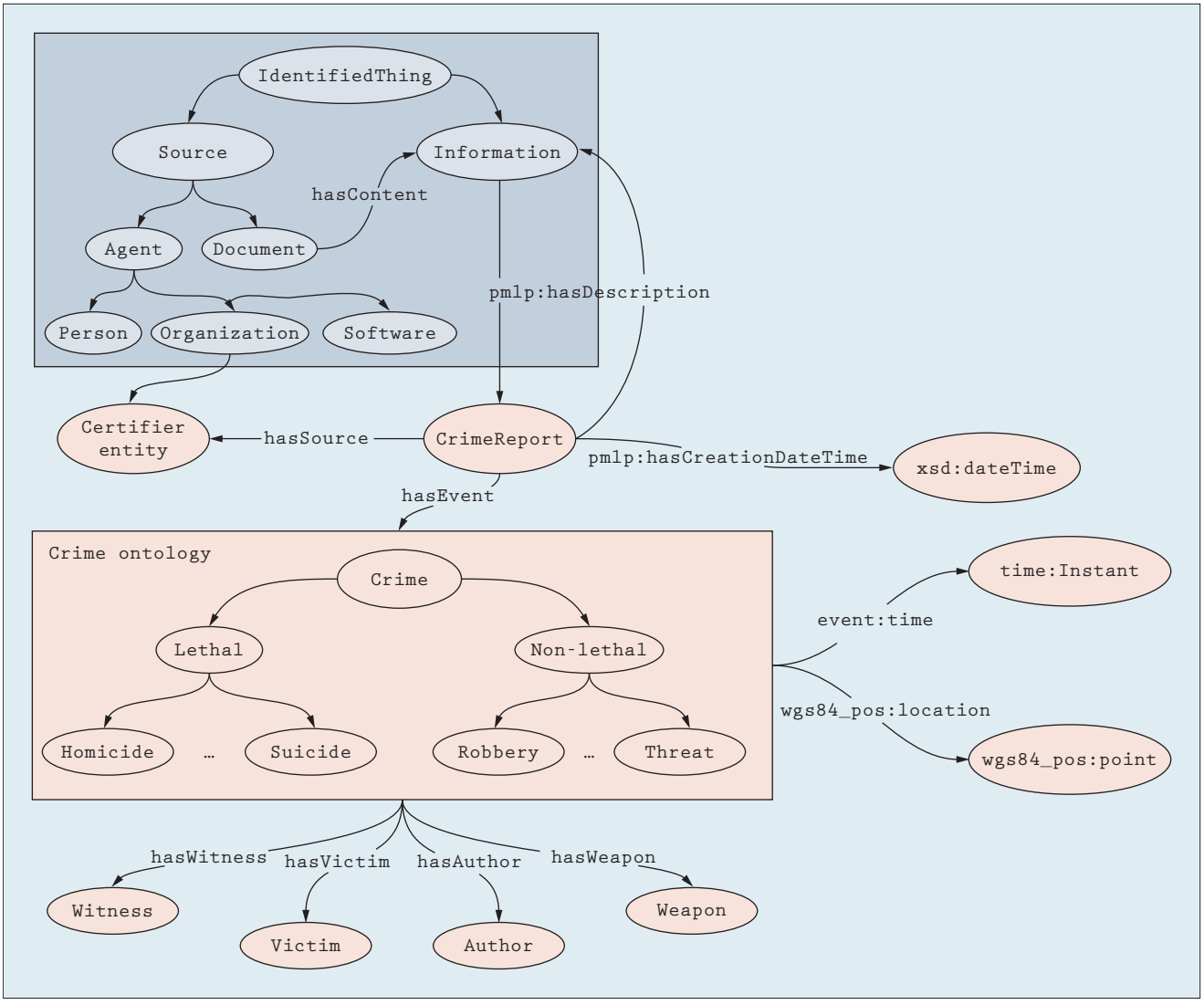

Figure 1. A piece of the crime and crime report ontologies and their relationship. At the core is the concept of crime with a categorization of lethal and nonlethal and the main components such as author, witness, victim, and weapon. The crime report ontology basically represents the source, the description, and the date of the report, establishing the link between concepts of the crime ontology with the provenance one.

agents they indicate, then to the ones those indicated trust, and so on.

Reputation and trust are updated based on interactions—namely, posting a crime, confirming a crime positively, confirming a crime negatively, and denouncing abuse. User interaction with the Information Layer indicates how much trust system users have in that information.

Mining Fake Reports

Reputation models, however, lack the level of granularity to capture malicious activities such as generation of a false trend that can come about with an excess of false reports. Identifying evidence of these problems through data mining is a possible approach.

We're investigating mining algorithms to identify patterns that indicate malicious activities by WikiCrimes users. In particular, we've tried to identify patterns that indicate abuses stemming from a specific group of individuals.[11] These malicious actions can't be captured solely through report analysis; they require a fuller investigation.

Our strategy is based on analyzing system users' social networks, specifically, the ones comprised of users who report the events that form a hot spot. The idea is to identify the existence of communities in the social network and recognize if one of them dominates event reporting. To do that, we resort to the community discovery algorithms so widely studied in social network analysis. Our community identification procedure

is also based on the removal of the weak edges (with respect to the number of shortest paths between all pairs that run along each edge of the graph).[12] However, for each iteration, we compare the percentage of the community size with the highest contribution to the hot spot formation and the percentage of this contribution with the respective values in the previous graph (without edge removal). The idea is to identify whether an increase in the utility justifies continuing the process of community refinement.

Another approach is to consider complex networks modeled from information obtained by users, reports, and locations where those reports were made.[13] Starting from a bipartite network model in which the vertices are individuals and census tracts, we projected a monopartite network of users in which the edges indicate the strength of connection between them. This connection strength indicates the degree of co-relatedness of crime reports made by two users in a particular place. Based on this modeling and on information such as the distribution of crime per census tract versus the distribution of reports from users, we found regularity within the context of WikiCrimes.

Specifically, when analyzing WikiCrimes data, we saw that Certifier Entities formed hubs, which makes sense because they have large amount of data and make their reports from various census tracts in the city. These hubs have an essential role in the behavior patterns of users and their reports of crimes, and seemed to be the key for detecting activities that might indicate fraud.

Automatically Updating WikiCrimes

News about crime brings out the characteristics, peculiarities, and interrelationships of the events and people involved and also helps us perceive trends that can improve public safety and feed WikiCrimes. Understanding the vast amount of information to determine requires NLP systems.

We defined an architecture for information extraction (IE) systems to explore descriptions of criminal occurrences to provide input for WikiCrimes. Specifically, we developed a software program called WikiCrimes Information Extractor (WikiCrimesIE) under that architecture,[14] which prescribes (in addition to the NLP module) user-oriented programming tools that improve interaction and manipulation of natural language Web content. The innovative part of this architecture lies in the semantic analysis module, which is based on SIM (Semantic Inferentialism Model)[15] and provides facilities for interacting with and extracting information from natural language texts that contain information of interest to public safety systems.

SIM proposes a new way to understand natural language in which semantic reasoning happens holistically, on top of pragmatic knowledge. Often, the information to be extracted from written texts is implicit, which requires drawing inferences from the use of concepts in the linguistic praxis. For instance, when we read the news, "John murdered his wife by shooting her to death after an argument at Solon Pinheiro Street," we can refute an assertion that the type of weapon used was a "white weapon" (a nonfirearm whose primary use is as a tool—a knife, needle, ax, or stick) and that the type of crime was "homicide." This is possible because we, as users of natural language, know the conditions in which the concepts "to shoot" and "to murder" can be used. For example, when using the concept "to murder" in the sentence "X murdered Y," we associate the crime committed by X with the death of Y. Such inferences don't come from the individual concepts "to murder" or "wife" but from their context in the sentence.

The use of SIM here is very important because information about crimes (type, weapon used, causes/motives) is often implicit in the journalistic text, and we must extract more complex inferences from linguistic practice.

Assisting the Map between Relational Data and the Crime Ontology

The definition of a language to be used as a pattern for opening data on criminal incidents is crucial but not enough to be effectively adopted by the community. Moreover, we must consider how user friendly the pattern is. Thus, it's essential that the correspondence between information represented in the pattern and information represented in police databases is easy to achieve. Elsewhere,[10] we described a method called D2R-Crime that seeks to accomplish this. It relies on two assumptions. First, because crime data is stored in relational databases, the Web publication thereof shouldn't require data replication. Second, the task of associating the original data with a standard (say, a crime ontology) shouldn't require learning another programming language.

To achieve the first requirement, we based our method to map relational data to RDF on systems that work on demand, where applications (typically, webservers) take requests from the Web and rewrite them as SQL queries. We chose to use an approach based on the D2R server because it's an open and free system for publishing

relational data on the Web.[16] It enables RDF and HTML browsers to navigate the content of non-RDF databases and lets applications query databases using SPARQL.

D2RCrime supports the publication of crime reports in RDF from relational databases. The goal is to help designers who don't have extensive knowledge in semantic technologies to map relational data into RDF. The crime ontology described earlier interactively guides the designer as he or she obtains the D2RCrime mapping between the ontology classes and the database tables with questions about how to retrieve tuples from the database that describe a particular class (or property). The aim is thus to use a language largely dominated by designers that clearly describes the concepts represented in the crime ontology.

To date, we've integrated D2RCrime into WikiCrimes, so that instances retrieved by WikiCrimes from a police department's relational databases via D2RCrime are plotted directly on a digital map.

WikiCrimes currently has more than 10,000 confirmed users. The project launched in January 2008 and since then has had more than 400,000 page visits from 186 countries. Most of the visitors come from Brazil, which makes sense, because we started it in Fortaleza, a city of 2.5 million inhabitants in Northeast Brazil. Accordingly, this hot spot has the most registered users and reported crime. Work through civic organizations, workshops, lectures, and a diverse and consistent campaign of local advertising were particularly intense in the city. Expansion to other regions is happening gradually, mainly through agreements with municipal governments of medium-sized cities and collaborators who maintain similarly themed blogs.

We believe that WikiCrimes, adopted on a larger scale, can become a complete tool for exploring the possibilities of citizen-centric applications. In many regions, the democratic experience is evolving from the bottom up, and WikiCrimes offers an easy way for citizens to acquire and share information about crimes and gauge how they react to services from their governments, participate in campaigns, and engage in civic life.

The lines of investigation we described here are neither exhaustive nor complete. They serve to show how the environment around WikiCrimes is rich and full of possibilities. Open issues persist and will drive our future research. Currently, we're investigating the possibility of generalizing the findings we uncovered in the WikiCrimes context to different domains. So far, we've built a platform for creating, maintaining, and hosting WikiMapps (www.wikimapps.com),[17] which is where we're investigating the possibility for map designers to characterize maps semantically. WikiMapps could make it possible to generate semantic crowd maps that have the power to create links to external sources that constitute useful and appropriate information in the map context. ■

Acknowledgments

The first author was partially funded by CNPq grant: 55977/2010 and 304347/2011.

References

1. M. Gillavry, "Collaborative Mapping and GIS: An Alternative Geographic Information Framework," *Collaborative Geographic Information Systems*, S. Balram and S. Dragicevic, eds., Idea Group Publishing, 2006, pp. 103–119.
2. J. Rouse, S.J. Bergeron, and T.M. Harris, "Participating in the Geospatial Web: Collaborative Mapping, Social Networks and Participatory GIS," *The Geospatial Web: How Geobrowsers, Social Software and the Web 2.0 Are Shaping the Network Society*, A. Scharl and K. Tochtermann, eds., Springer, 2007, pp. 153–158.
3. D. Lathrop and L. Ruma, *Open Government: Collaboration, Transparency, and Participation in Practice*, O'Reilly Media, 2010.
4. T. Kahn, "Boletim de Ocorrência: Prover Para Poder Prever," *Forum Brasileiro de Segurança Pública*, 2008; www.forumseguranca.org.br/artigos/.
5. V. Furtado et al., "Collective Intelligence in the Law Enforcement: The WikiCrimes System," *Information Science*, vol. 180, no. 1, 2010, pp. 4–17.
6. P.J. Brantingham and P.L. Brantingham, *Environmental Criminology*, Waveland Press, 1991.
7. J.R. Hobbs and F. Pan, "An Ontology of Time for the Semantic Web," *ACM Trans. Asian Language Information Processing*, vol. 3, no. 1, 2004, pp. 66–85.
8. Y. Raimond and S.A. Abdallah, "The Event Ontology," 2006; http://purl.org/NET/c4dm/event.owl.
9. D. McGuinness et al., "PML2: A Modular Explanation Interlingua," *Proc. AAAI 2007 Workshop on Explanation-Aware Computing*, AAAI, 2007, pp. 22–23.
10. J. Tavares, V. Furtado, and H. Santos, "Open Government in Law Enforcement: Assisting the publication of Crime Occurrences in RDF from Relational Data," Presentation at the AAAI Fall Symp. Open Government Knowledge: AI Opportunities and Challenges, AAAI, 2011; www.cs.umbc.edu/~finin/noindex/aaai-2011-fss-ogk/presentations/ogk2011_submission_6_crime_vasco.pptx.
11. V. Furtado et al., "A Method for Identifying Malicious Activity in

Collaborative Systems with Maps," *Proc. Int'l Conf. Advances in Social Network Analysis and Mining* (ASONAM), ACM, 2009, pp. 334–337.
12. M.E.J. Newman and M. Girvan, "Finding and Evaluating Community Structure in Networks," *Physical Rev. E*, vol. 69, no. 2, 2004, article 026113.
13. C. Caminha and V. Furtado, "Modeling User Reports in Crowd Maps as a Complex Network," *Proc. WebScience Track in WWW2012*, IEEE CS, 2012; http://www2012.wwwconference.org/proceedings/nocompanion/wwwwebsci2012_caminha.pdf.
14. V. Pinheiro et al., "Natural Language Processing Based on Semantic Inferentialism for Crime Information Extraction," *Proc. IEEE Intelligence and Security Informatics*, IEEE, 2010, pp. 19–24.
15. V. Pinheiro et al., "Information Extraction from Text Based on Semantic Inferentialism," *Proc. 8th Int'l Conf. Flexible Query and Answering Systems* (FQAS 2009), LNAI 5822, Springer, T. Andreasen et al., eds., 2009, pp. 333–344.
16. C. Bizer and R. Cyganiak, "D2R Server: Publishing Relational Databases on the Semantic Web," Poster at the 5th Int'l Semantic Web Conf., 2006; http://richard.cyganiak.de/2008/papers/d2r-server-iswc2006.pdf.
17. H. Santos and V. Furtado, "A Service-Oriented Architecture for Assisting the Authoring of Semantic Crowd Maps," to be published in *Proc. Brazilian Symposium of Artificial Intelligence*, Springer, 2012.

Vasco Furtado is a computer science professor at the University of Fortaleza and a researcher at ETICE, Brazil. His research interests include Web science, data mining, and e-government. Furtado has a PhD in computer science from the Université d'Aix Marseille III. Contact him at vasco@unifor.br.

Carlos Caminha is a graduate student at the University of Fortaleza and a system analyst at Wikinova. His research interests include data mining, complex networks, and human-computer interaction. Caminha has an MS in computer science from the University of Fortaleza. Contact him at Carlos.o.c.neto@gmail.com.

Leonardo Ayres is a systems analyst at Wikinova in Brazil. His research interests include the Semantic Web, distributed systems, and crowd mapping. Ayres has an MS in computer science from the University of Fortaleza. Contact him at leonardo@wikinova.com.br.

Henrique Santos is a graduate student at the University of Fortaleza. His research interests include the Semantic Web, distributed systems, and crowd mapping. Santos has an MS in computer science from the University of Fortaleza. Contact him at hensantos@gmail.com.