Journal of Computer Science & Technology, Volume 17, Number 2, October 2017

- ORIGINAL ARTICLE -

An Expert-driven Ontology-based Approach to the Collaborative Acquisition of Information

Alejandro Sánchez[1], Leo Ordinez[1], Sergio Firmenich[1], Damián Barry[1], and Rodrigo Santos[2]

[1]*Laboratorio de Investigación en Informática (LINVI), Dep. de Informática, UNPSJB, Puerto Madryn, Argentina*
aljsanchez@gmail.com, leo.ordinez@gmail.com, sergio.firmenich@lifia.info.unlp.edu.ar, damian_barry@unpata.edu.ar
[2]*Dep. de Ingeniería Eléctrica y Computadoras, IIIE UNS-CONICET, Bahia Blanca, Argentina*
ierms@criba.edu.ar

Abstract

This article presents design decisions for an ontology-based framework supporting an expert-driven approach to the collaborative acquisition, integration and analysis of information. It describes a three-tier organization in which components facilitate establishing an information system to assist domain experts in addressing multi-causal dynamic situations where heterogeneous information sources must be integrated and the interaction of a variety of actors must be supported in large geographical areas with possibly low or intermittent connectivity levels. The approach and envisioned framework are illustrated with a case-study in the design of bikeways in urban zones.

Keywords: expert-driven, ontology, collaborative acquisition of information

1 Introduction

Domain experts are key in the success of an organization. They typically rely on information systems to decide on core issues under time pressure. However, it is often the case that information systems fail to provide enough information: either because a suitable query is missing, or because data repositories lack the sough information. This inadequacy might prevent, for example, a business opportunity to be analyzed, or even recognized.

In some contexts, the sough information is hard to obtain. The reason lies in the complex nature of the reality from which it must be collected. We refer to multi-causal dynamic situations where a variety of actors interact in large geographical areas with possibly low or intermittent connectivity levels. Consider, for instance, an information system for assisting in the design of bikeways (exclusive tracks, paths or lanes for cyclists) in an urban zone, where the actual behavior of cyclist must be learned. A sensible solution in such cases is to use applications for the collaborative acquisition of information, where the end-users – which we call *data collectors* – provide the required information themselves.

IT-experts play the role of developing and modifying information systems to satisfy domain expert needs. They define new queries to address their needs, or they implement applications for the collaborative acquisition of information that is otherwise unavailable.

Unfortunately, the process of developing or modifying information systems in organizations is typically complex and takes longer than convenient. The conceptualization gap between domain- and IT-experts makes difficult defining requested queries right in one attempt. This process usually requires several cycles of interaction. In addition, the task is often complex, as accessing information might require integrating multiple and heterogeneous repositories in the organization. In the case of the implementation of an application for the collaborative acquisition of information, the number and duration of development cycles entails a more significant cost. These cycles might cause the loss of business opportunities or even discourage domain-experts from analyzing them.

We propose an approach to address this type of situations. It refines the work presented in [1], where a set of requirements for an expert-driven framework are presented. In this article, a framework is envisioned to address these requirements. We propose an ontology-based [2] solution with components adopting a three-tier organization.

The expert-driven quality of the framework comes from its aim: allowing domain-experts address their information needs with a minimum of intervention of IT-experts, reducing required interaction cycles. The envisioned framework will allow domain-experts to visually specify queries and generate applications for the collaborative acquisition of information.

Citation: A. Sánchez, L. Ordinez, S. Firmenich, D. Barry and R. Santos. *"An Expert-driven Ontology-based Approach to the Collaborative Acquisition of Information"*. Journal of Computer Science & Technology, vol. 17, no. 2, pp. 127–134, 2017.

Received: March 01, 2017. **Revised:** June 21, 2017. **Accepted:** June 26, 2017.

The approach is ontology-based in multiple ways. Domain-experts visually specify and execute queries upon an ontology that approximates the conceptualization they use in their work. This ontology is also used in the integration of the relevant information sources in the organization. When domain-experts define applications for the collaborative acquisition of information, they use a visual tool that allows indicating elements of the ontology as data sources (or sinks) for visual controls in the generated application. Thus, information collected by generated applications is by design already mapped to the ontology that domain-experts are familiarized with.

We illustrate the approach with a case study in the design of bikeways in urban zones. Bikeways must meet a given set of criteria to be successfully adopted by citizens. For instance, they must: be built upon routes that are actually taken by cyclists; link central places in the city (*e.g.*. schools), visit tire repair shops, avoid routes for accessing hospitals – which must be free to let ambulances a swift passage, and elude streets used for pluvial drainage – since they might become difficult to use during rainy days. An information system for the design of bikeways must collect information about the behavior of cyclists, integrate it with a number of heterogeneous information sources, and answer the questions that domain-experts pose to ensure that criteria is met by an eventual bikeway. The implementation of this information system is currently under development.

The rest of the article is organized as follows: section 2 provides an overview of related work; section 3 describes the bikeways case study; section 4 presents the expert-driven foundations of the approach; section 5 exposes the fundamental role of ontologies; section 6 proposes a three-tier organization for the framework supporting the approach; and section 7 summarizes the article and describes next steps.

2 Related Work

Research related to the approach and envisioned framework described in this article can be grouped in two major areas: *end-user development* and *ontology-based access to information*.

The research area of end-user development (or programming) aims at providing the ability of creating and modifying software artifacts by end-users lacking any programming background [3]. In [4], the author provides a review of end-user development efforts, including end-user development for web applications, and end-user development for mobile applications. With a more broad focus, the authors coin the term *end-user software engineering* in [5] to group the activities (for instance, requirements elicitation, specification, design, reuse, integration) carried out by end-users to develop the software they require. They classify these activities according to their contribution in the lifecycle and discuss technologies supporting them. Among the results, they mention that the end-user approach is extending to web platforms and mobile devices, and that research is focusing more on application domains rather than in programming paradigms. In [6] the authors discuss the state of the art of end-user software engineering, highlighting the need of avoiding silos and thus fostering greater levels of collaboration among end-user initiatives.

Research on ontology-based integration of information goes decades back [7]. A survey of early approaches can be found in [8]. A modern and consolidated approach, called *ontology-based data access (ODBA)*, allows defining queries using an ontology formalizing the concepts the user is familiar with, and then translating such query into optimized queries to be run over the different involved information sources. A number of relevant works in this area can be found in [9]. Tools supporting this approach include [10] and [11].

Research on end-user development for accessing information sources also goes decades back. Consider, for instance, the survey documented in [12], where techniques for the visual database query formulation are studied. Ontologies are used in [13] to facilitate visual query formulations. Focusing on usability, in [14], the authors describe an ontology-based visual query system for domain-experts in the context of a case-study.

3 Bikeways design in urban zones

City governments around the world are encouraging biking. Consider, for instance, initiatives such as [15]. Biking constitutes an economic and healthy transportation alternative with the extra benefit (among others) of producing zero polution.

However, cities are seldom found to be cycling friendly. The major problem is that they usually lack bikeways. A city that decides to acquire this quality faces a complex task. The proper design of bikeways requires analysing information coming from heterogeneous sources, and acquiring information that is hard to collect. Heterogeneous information sources include, for example, city streets maps, traffic flow, pluvial drainage, public transportation usage, and topographic maps. Information that is typically unavailable and must be acquired concerns the behaviour of cyclists; the actual routes they prefer and at which rates they use them.

Experts in urban design use this information to ensure that bikeways are properly defined according to a given set of criteria, which might include rules as enumerated below.

- Actual routes taken by cyclist must be used unless another rule indicates the opposite.
- The most relevant spots in the city – called centralities, or their surroundings, must be visited.

- Wider streets must be preferred.
- Street segments where tire repair shops are already located must be preferred.
- Bikeways must avoid:
  - access paths to hospitals and fire stations (and similar emergency/security buildings),
  - high levels of motorized traffic,
  - access to parking lots,
  - pluvial drainage, and
  - steep segments of streets.

Information systems must be developed to assist experts in the definition of bikeways that met these criteria. We use this problem to illustrate how the proposed approach can be applied. We argue it will facilitate the integration of heterogeneous information sources, an expert-driven definition of queries, and an expert-driven development of applications to collaboratively acquire otherwise hard to collect information.

4 An expert-driven approach

Domain experts define and modify information system themselves using the framework that the approach envisions. Components and workflow supporting the approach must attend the requierments enumerated below.

R1. Domain experts define themselves the information they need and the way in which this information is automatically converted into displayable information.

R2. Domain experts specify themselves applications for the collaborative acquisition of information (in mobile devices) from data collectors, which are automatically derived.

R3. Changes in definitions specified by domain experts are automatically propagated to derived applications and components for displaying information.

R4. Information sources are integrated and managed in a format-agnostic way. They comprise repositories that
   (a) are found across organizations,
   (b) store data from sensors, and
   (c) collect data obtained from *Data Collectors* using applications for the collaborative acquisition of information, which are possibly running on mobile devices.

R5. Information gathering by derived applications continues despite poor or intermittent levels of connectivity.

R6. Data collectors are rewarded. Consider, for instance, an scenario in which data collectors collect information while participating in a game (*gamification*).

These requirements characterize an expert-driven approach to address information needs of domain experts in multi-causal dynamic situations where a variety of actors interact in large geographical areas with possibly low or intermittent connectivity levels. Requirements R1 to R3 focus on support allowing domain experts to obtain the information they need without engaging in development cycles with IT-experts. Requirement R3 emphasises the dynamic nature of situations. Connectivity issues are considered by R5 and involvement of data collectors is sought by R6.

Returning to our case-study, the framework the requirements envision facilitates the work of domain experts in the design of bikeways as follows: urban design experts will define the queries they need to ensure that a bikeway satisfies an imposed criteria – R1 and R3; urban design experts will define an application that cyclist will use in their mobile devices to collect information about the paths they follow and with which frequency – R2; whenever the cyclist is disconnected the application will continue storing information about his behaviour – R5; cyclist will use the application since it informs them about their performance, and it might also be a game with a ranking and prizes for the best ranked – R6.

5 An ontology-based approach

Part of the requirements for the envisioned framework are addressed relying on ontologies. The framework will include visual tools with an interface prepared to intuitively interact in terms that domain experts are familiar with. These terms are taken from an ontology that formalizes the conceptualization they share, *i.e.*, an ontology of their domain. This ontology serves multiple purposes as we describe next.

The ontology is used to integrate information coming from heterogeneous repositories. We refer to this ontology as the *global ontology* in the sequel. Figure 1 depicts a view highlighting the role of the global ontology within the approach. Each repository has an associated ontology – called *local ontology* – which is developed to import the relevant information from it. Then, a number of mappings defined between the global ontology and the local ontologies, following a star-like configuration, support an integrated view of the multiple repositories. This design decision constitutes a format-agnostic solution, and thus satisfies requirement R4.

The adoption of this design decision in the bikeways case-study requires a number of ontologies. A global ontology, which we call the *bikeway design ontology*, and local ontologies as follows: a *city streets ontology*, a *traffic flow ontology*, a *pluvial drainage ontology*, a *public transportation usage ontology*, and a *topographic ontology*. It also requires defining mappings between the bikeway design ontology and each of the local ontologies.

Ontologies are developed by IT-experts and domain experts in collaboration. The global ontology must

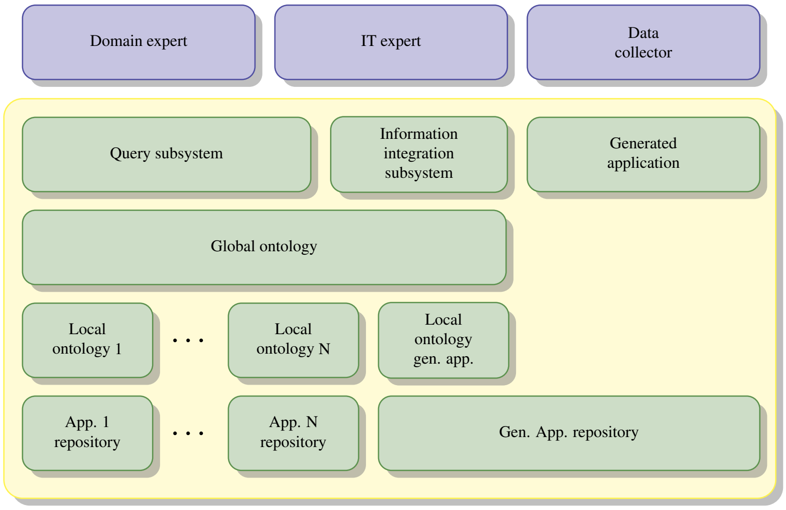

Figure 1: Ontology-based approach

faithfully capture the conceptualization that domain experts require. An ontology development methodology that suits the needs of the specific context can be chosen to organize this collaborative development (for instance, methodologies described in [16]). IT-experts develop a local ontology for each of the repositories that includes relevant data. They might use tools such as [11] to facilitate their work. IT- and domain-experts jointly develop the mappings between the global and local ontologies.

The global ontology also serves as basis for specifying and executing queries. A visual intuitive interface allows domain experts to specify queries over the global ontology. This query is then translated to be executed in local ontologies or repositories. The results are collected and consolidated back into the global ontology. Domain experts are then notified to access the results.

The generation of applications for the acquisition of information relies on the ontology as well. A visual interface allows domain experts to customize and combine generic widgets from which an application is generated. Part of this customization consists in indicating sources (and sinks) of information for this generic widgets. These sources are initially indicated in terms of the global ontology. Concepts that are absent are specifically introduced and remain in the local repository only. When the application is generated, a local ontology and mappings to the global ontology are automatically generated. After its deployment, the generated application is operated by the data collector who enters the otherwise missing information to the application's repository. This information becomes immediately available for executing queries through the local ontologies and mappings.

6 A three-tier organization

The envisioned framework adopts a three-tier organization. An overview showing the major components is depicted in Figure 2. Components in each tier address one of two major concerns: collaborative acquisition of information (in green), and integration and analysis of information (in orange). They participate in subsystems, grouped by concern, as detailed below.

- Collaborative acquisition of information
  - Application generation
  - Information acquisition
  - Template development
- Integration and analysis of information
  - Query specification and execution
  - Ontology management

The next paragraphs detail how these subsystems address the requirements proposed in Section 4.

6.1 Application generation

This subsystem aims at allowing domain experts define themselves applications for the collaborative acquisition of information. It corresponds to requirements R2, R3 and R6, and comprises the *dashboard application*, the *application generator*, and the *application repository*.

The dashboard application adopts the metaphor in which domain experts use dashboards to configure the data they expect to be collected by data collectors. Once defined, domain-experts generate and share the mobile applications, which are automatically updated if the specifications in dashboard are modified. The application generator provides an API for accessing to application templates and widgets, to a version control system, and to a build and deployment system. The application repository stores generated applications in their different stages.

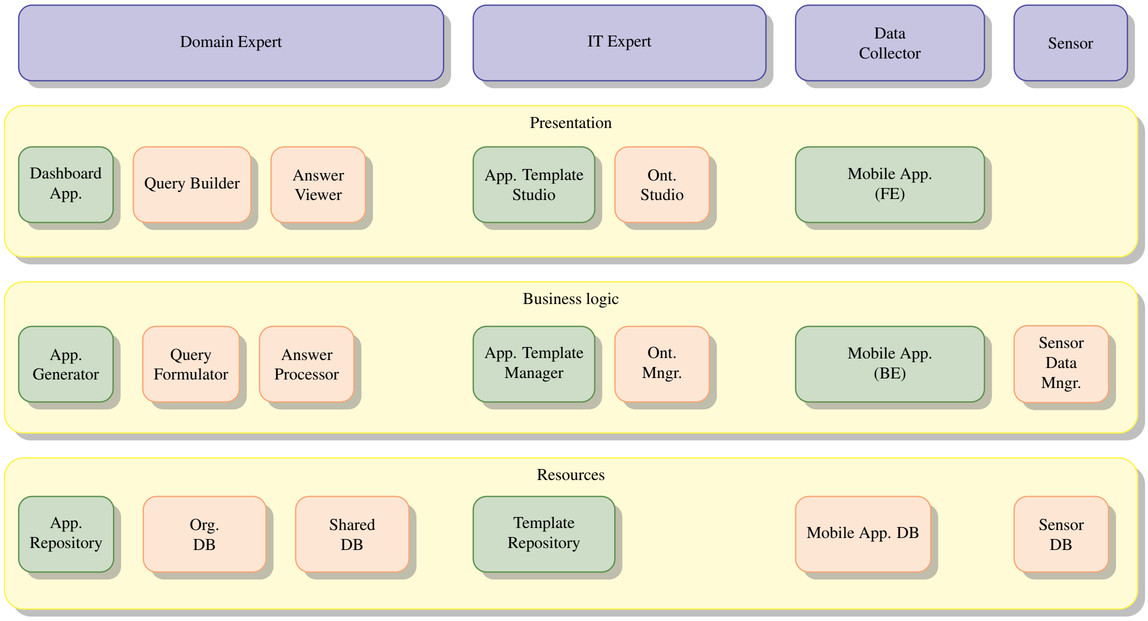

Figure 2: Three-tier organization

A prototype of the dashboard application was developed (see Figure 3). In its present form, it is a web application that allows domain experts to drag and drop elements from a widgets pallet to the dashboard definition form. From this form, the domain expert may configure each piece of data to be collected for each particular widget, configure widgets name, description, data validation, etc.

The libraries that underpin the dashboard application are prepared to easily develop new kind of widgets. The main idea is that new widgets could be plugged in the application and automatically be included in existing dashboards. These new widgets must have associated a specification of the kind of field to be used (map, input for integer, etc.) in order to generate the mobile application

The dashboard application provides an extension point for adopting a gamification approach. Domain-experts may define a schema by specifying a particular kind of reward strategy, and the corresponding reward for each data collector collaborating with the system. Future prototypes of the dashboard application are expected to include this feature, and thus satisfy requirement R6. They are also expected to use the global ontology for indicating sources (and sinks) of information, and to include widgets for an off-line mode of work. The latter to cope with requirement R5.

Urban design experts use the dashboard to define applications for acquiring data. They combine widgets indicating the sources (and sinks) of information, including concepts from the bikeways ontology, or sensors in mobile devices. They generate the application when it is completely specified.

An example of use of this application is shown in Figure 3, where an expert defines a dashboard where people may collaborate by reporting avoidable points for bike pathways. In this simple example, the experts has defined a dashboard where users may send the kind of avoidable point, the location (which is set to automatic) and also asks collaborators to take a picture. At the right, the a menu is offered to add more form items to the collaboration form.

6.2 Information acquisition

The information acquisition subsystem comprises the *mobile application front-end*, the *mobile applicaton back-end*, the *mobile application database*, the *sensor data manager*, and the *sensor database*. The first three are generated by the application generation subsystem described in section 6.1.

Data collectors start using generated application once they are released through pre-arranged sites, possibly engaged within the context of a game. Data is stored in a generated database that is already mapped to the global ontology, as it is described in section 6.5. Then, it is also available to be consulted by domain-experts, as it is further explained in section 6.4.

The sensor data manager and sensor database provides support for cases in which no mobile application is needed, and data is directly collected from sensor. In such cases, IT-experts configure sensor to send data to the sensor manager.

Coming back to our case study, cyclists find out that the Municipality of their city is launching a weekly contest for citizens. They are supposed to use a mobile application to measure their cycling activities, and the best ranked citizen win prizes that range from free bike reparations to new bikes. Citizens become users of the generated application, and then they become data collectors in the collaborative acquisition of information about their cycling behaviour.

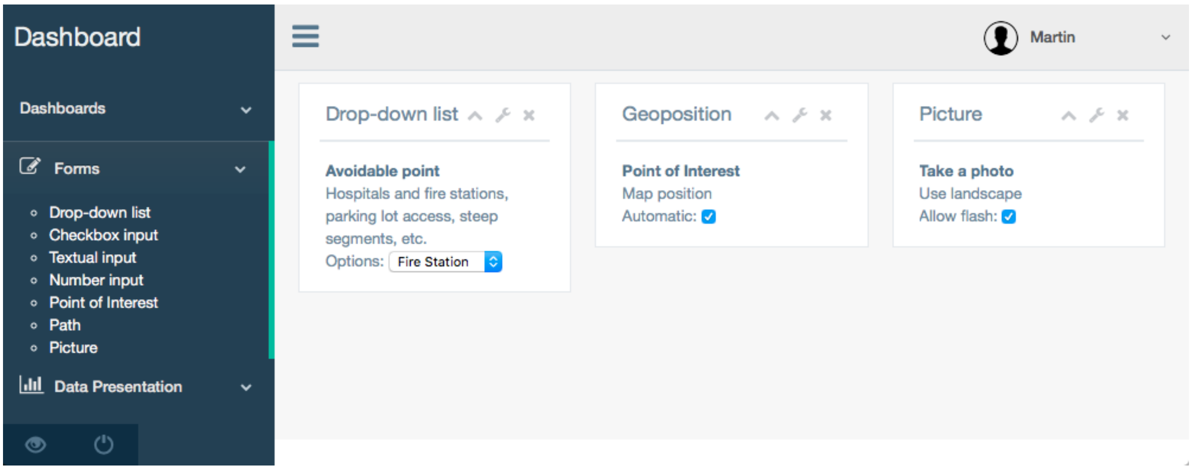

Figure 3: Dashboard screen capture

6.3 Template development

The template development subsystem supports the provision of new application templates and widgets by IT-experts. Once new application templates and widgets are completed, they are made available to the dashboard application (described in section 6.1). In other words, it allows creating and maintaining the building blocks used to satisfy requirements R2 and R3. It comprises the *application template studio*, the *application template manager* and the *template repository*.

IT-experts use the subsystem to develop a variety of application templates and widgets. Application templates that provide a basis for games are included. Widgets that involve processing and displaying information can also be linked to elements in the global ontology, to be used as sources (or sinks) of information.

In the context of the bikeways case-study, widgets that support street maps and GPS sensors are required. These widgets require a back-end where GPS information is translated into trajectories described in terms of street segments. Alternative, widgets supporting heart rate monitors and other body parameters could also be included.

6.4 Query specification and execution

This subsystem allows domain-experts to specify, execute and view the results of their queries. The susbsystem mainly addresses requirement R1. It comprises the *query builder*, the *query formulator*, the *answer viewer*, the *answer processor*, *organization databases*, *shared databases*, *mobile application databases*, and *sensor databases*.

Domain-experts visually specify a query in terms of the global ontology using the query builder. The query formulator translates the query to local ontologies according to the defined mappings, it optimizes the resulting queries and distributes their execution. The answer processor collects responses from shared databases and translates them back to the global ontology. Domain experts are notified when the answer processor finishes, and then are able to visualize the results with the answer viewer.

This subsystem offers predefined queries for visualizing the information collected with generated applications. When a generated application is released, the specification of the corresponding queries are included.

Urban design experts use the query builder to specify queries to understand which street segments are candidates to be used for bikeways.

6.5 Ontology management

The ontology management subsystem is mainly concerned with the integration of the different information sources, *i.e.*, addressing requirement R4. It comprises the *ontology studio* and the *ontology manager*.

IT-experts create and maintain ontologies and mappings with the ontology management subsystem. The ontology studio allows them to visually specify the global ontology, local ontologies, and mappings. Domain-experts collaborate with them in this task. The ontology manager provides a back-end support for the ontology studio. It provides an API for storing and modifying ontologies and mapping in repositories that are not shown in Figure 2. Well-known ontologies might be used when defining global and local ontologies. IT-experts might also indicate where to store and execute queries upon a given ontology, since it might require specific feature-support.

The bikeways design ontology, the local ontologies (city streets ontology, traffic flow ontology, . . .), and the mappings among them, are developed and maintained using the ontology studio. In addition standards for representing spatial and temporal information, such as GeoSPARQL [17] and SOWL [18] are considered, as long as the repositories supporting them such as

[19].

7 Conclusions

This article describes design decisions for a framework supporting an expert-driven approach to the collaborative acquisition, integration and analysis of information. It allows domain-experts to visually specify their information needs themselves, either through queries, when the information is known to exists in repositories, or through the specification of application for the collaborative acquisition of otherwise non-available information. The requirements (R1 to R6) are shown to be addressed by the proposed design decisions. Both the approach and the envisioned framework are illustrated with a case-study in the design of bikeways in urban zones.

Future work includes addressing challenges as follows: integrating and analysing information with specific attributes such as spatial and temporal; providing the dashboard application with support of gamification strategies; and using ontologies to indicate information sources (and sinks) for widgets during the specification of applications. Currently ongoing work includes the implementation of the bikeways information system, which is the next expected concrete result.

Competing interests

The authors have declared that no competing interests exist.

References

[1] L. Ordinez, R. Santos, D. Barry, S. Firmenich, and A. Sánchez, "An expert-driven approach for collaborative knowledge acquisition," in *Anales del Congreso Argentino de Ciencias de la Computación, CACIC 2016* (U. N. de San Luis, ed.), pp. 986–995, October 2016.

[2] T. R. Gruber, "A translation approach to portable ontology specifications," *Knowl. Acquis.*, vol. 5, pp. 199–220, June 1993.

[3] H. Lieberman, F. Paterno, and V. Wulf, eds., *End User Development*. Springer, 2006.

[4] F. Paternò, "End user development: Survey of an emerging field for empowering people," *ISRN Software Engineering*, vol. 5, no. 11, pp. 1–11, 2013.

[5] A. J. Ko, R. Abraham, L. Beckwith, A. Blackwell, M. Burnett, M. Erwig, C. Scaffidi, J. Lawrance, H. Lieberman, B. Myers, M. B. Rosson, G. Rothermel, M. Shaw, and S. Wiedenbeck, "The state of the art in end-user software engineering," *ACM Comput. Surv.*, vol. 43, pp. 21:1–21:44, Apr. 2011.

[6] M. M. Burnett and B. A. Myers, "Future of end-user software engineering: Beyond the silos," in *Proceedings of the on Future of Software Engineering*, FOSE 2014, (New York, NY, USA), pp. 201–211, ACM, 2014.

[7] G. Wiederhold, "Intelligent integration of information," *SIGMOD Rec.*, vol. 22, pp. 434–437, June 1993.

[8] H. Wache, T. Voegele, T. Visser, H. Stuckenschmidt, H. Schuster, G. Neumann, and S. Huebner, "Ontology-based integration of information - a survey of existing approaches," in *IJCAI-01 Workshop: Ontologies and Information* (H. Stuckenschmidt, ed.), pp. 108–117, 2001.

[9] D.-E. Spanos, P. Stavrou, and N. Mitrou, "Bringing relational databases into the semantic web: A survey," *Semant. web*, vol. 3, pp. 169–209, Apr. 2012.

[10] D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, A. Poggi, M. Rodriguez-Muro, R. Rosati, M. Ruzzi, and D. F. Savo, "The mastro system for ontology-based data access," *Semantic Web Journal*, vol. 2, no. 1, pp. 43–53, 2011. Listed among the 5 most cited papers in the first five years of the Semantic Web Journal.

[11] M. Rodriguez-Muro and D. Calvanese, "Quest, an OWL 2 QL reasoner for ontology-based data access," in *Proceedings of OWL: Experiences and Directions Workshop 2012, Heraklion, Crete, Greece, May 27-28, 2012* (P. Klinov and M. Horridge, eds.), vol. 849 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2012.

[12] T. Catarci, M. F. Costabile, S. Levialdi, and C. Batini, "Visual query systems for databases," *J. Vis. Lang. Comput.*, vol. 8, pp. 215–260, Apr. 1997.

[13] M. Giese, D. Calvanese, P. Haase, I. Horrocks, Y. Ioannidis, H. Kllapi, M. Koubarakis, M. Lenzerini, R. Müller, M. Rodriguez-Muro, z. Özcep, R. Rosati, R. Schlatte, M. Schmidt, A. Soylu, and A. Waaler, "Scalable end-user access to big data," in *Big Data Computing* (R. Akerkar, ed.), CRC Press, 2013.

[14] A. Soylu, M. Giese, R. Schlatte, E. Jimenez-Ruiz, O. Ozcep, and S. Brandt, "Domain experts surfing on stream sensor data over ontologies," in *Proceedings of the 1st Workshop on Semantic Web Technologies for Mobile and Pervasive Environments (SEMPER@ESWC 2016)*, vol. 1588, (Heraklion, Greece), CEUR-WS.org, 2016.

[15] J. Dekoster and U. Schollaert, *cycling: the way ahead for towns and cities*. boulevard du Triomphe 174; B-1160 Bruxelles.: European Commision, 1 ed., 1999.

[16] A. Gómez-Pérez, M. Fernández-López, and O. Corcho, *Ontological Engineering: With Examples from the Areas of Knowledge Management, e-Commerce and the Semantic Web. (Advanced Information and Knowledge Processing)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2007.

[17] Open Geospatial Consortium, "Geosparql - a geographic query language for rdf data - v1.0." Available at: `http://www.opengeospatial.org/standards/geosparql`, 2011. Accessed on 2017-06-30.

[18] K. Stravoskoufos, E. G. M. Petrakis, N. Mainas, S. Batsakis, and V. Samoladas, "Sowl ql: Querying spatio-temporal ontologies in owl," *Journal on Data Semantics*, vol. 5, no. 4, pp. 249–269, 2016.

[19] "Parliament (tm) - a high-performance triple store, sparql endpoint, and reasoner." Available at: `http://parliament.semwebcentral.org`. Accessed on 2017-06-30.