Exploiting Smart City Ontology and Citizens' Profiles for Urban Data Exploration

Devis Bianchini[(✉)], Valeria De Antonellis, Massimiliano Garda, and Michele Melchiori

Department of Information Engineering, University of Brescia, Via Branze 38, 25123 Brescia, Italy
{devis.bianchini,valeria.deantonellis, m.garda001,michele.melchiori}@unibs.it

Abstract. Smart Cities are complex systems, collecting together huge amounts of heterogeneous data mainly concerning energy consumption, garbage collection, level of pollution, citizens' safety and security. In the recent years, several approaches have been defined to enable Public Administration (PA), utility and energy providers, as well as citizens, to share and use information in order to take decisions about their daily life in Smart Cities. Research challenges concern the study of advanced techniques and tools to enable effective urban data exploration. In this paper, we describe a framework that combines ontology-based techniques and citizens' profiles in order to enable personalised exploration of urban data. Ontologies may provide a powerful tool for semantics-enabled exploration of data, by exploiting the knowledge structure in terms of concepts organised through hierarchies and semantic relationships. Smart City indicators are used to aggregate data that can have different relevance for target users, the activities they are performing and their role (e.g., PA, utility and energy providers, citizens) within the Smart City. Ontologies combined with users' profiles enable effective and personalised recommendation and exploration of urban data.

Keywords: Urban data exploration · Smart city indicators
Ontologies · Smart cities

1 Introduction

In a Smart City, investments in human and social capital and modern ICT infrastructure fuel sustainable economic growth and a high quality of life, with a wise management of natural resources, through citizens' participation [7]. In this context, techniques and tools for sharing and exploring urban data about energy consumption, garbage collection, level of pollution, citizens' safety and security are gaining momentum to enable the citizens to take decisions about their daily

life, depending on different roles they play in the Smart City. Public Administration (PA), utility and energy providers may need to explore urban data at city level, in order to take actions for improving citizens' life. Building managers can use urban data concerning the administered buildings to take decisions for their daily activities (e.g., they can inspect information on energy consumption in order to implement energy saving actions). Citizens need new tools to explore open urban data about pollution, garbage collection, safety and security to take autonomously decisions on activities they may perform or to compare their own energy consumption against average values at building, district and city level to take virtuous behaviours. Let's consider two data exploration scenarios focused on different types of users. John is the manager of several buildings located in different districts of the Smart City. John monitors electrical consumptions of the buildings, in order to implement energy saving policies (e.g., introduction of LED lamps in common spaces or planning renovation work to increase the energy efficiency class). Alice is a citizen who is enthusiastic about bicycle and wants to inspect whether environmental status (in terms of pollution and air quality) is suitable for practising outdoor leisure activities. Challenging issues are related to the capability of John and Alice to fruitfully exploit available information.

In this context, we propose an ontology-based framework for enabling personalised exploration of urban data. In the framework, different categories of end users can be defined and their profiles are taken into account to suggest fruitful exploration, proactively supporting users for making decisions according to their interests. Through the framework, users with different roles may share, access and possibly update urban data coming from different types of data sources, guided by the knowledge structure provided by the ontology and according to their own interests. The framework supports the exploration of indicators that aggregate urban data from various data sources about energy consumption, garbage collection, level of pollution, citizens' safety and security. Actually, indicators provide a comprehensive view over underlying data according to several perspectives without being overwhelmed by the data volume [13]. Indicators are computed at the levels of building, district or city, aggregating data that can have different relevance with respect to the various target users, also considering the activities they perform and their role in the Smart City. Our framework relies on both a Smart City Ontology (SCO), which provides a powerful tool for semantics-enabled exploration of urban data, and on information associated with users' profiles. On the one hand, the SCO is used to properly represent indicators in terms of concepts, hierarchies and semantic relationships. The resulting knowledge organisation can be used for the selection and exploration of the most suitable indicators for a specific user and request. Moreover, inheritance in the ontology hierarchies allows exploration at different granularity levels. On the other hand, users' profiles are exploited to refine the list of suggested indicators. With reference to the examples given above, the framework may allow John to monitor electrical consumption of administered buildings, by exploiting the indicators hierarchy in the ontology to distinguish electrical consumption according to different perspectives (e.g., consumption in common

spaces, consumption of elevators), and to compare average values of consumption with other buildings at district or city level. Furthermore, Alice may use the framework to make decisions about her activities, by observing specific indicators (e.g., to avoid sport activities when pollution levels overtake tolerance thresholds).

Novel contribution of the framework relies on the two-fold nature of the approach: a recommendation step, which selects candidate indicators, is followed by and interleaved with an interactive exploration step, which permits to refine the set of indicators. These steps are based on the Smart City Ontology and users' profiles in a combined way. Firstly, indicators are selected by filtering available indicators, exploiting the information in the request issued by the user, as well as his/her profile. Once a set of candidate indicators has been identified, semantics-enabled data exploration is enabled, where concepts hierarchies and semantic relationships in the ontology are used to further refine and rank the indicators of interest. A preliminary validation of the framework has been performed in the context of the Brescia Smart Living (BSL) Italian project[1], which promotes a holistic view of the city where different types of data must be collected and properly explored to provide new services to several city stakeholders and, in particular, educational indications to citizens in order to promote virtuous behaviours.

The paper extends the research presented in [10] with the following innovative contributions: (i) a formalisation of the candidate indicators selection algorithm; (ii) a ranking function, applied to the set of candidate indicators, and (iii) a refinement of the framework architecture, providing additional details about the services and their constituent modules, within each layer.

This paper is organized as follows: in Sect. 2 we compare our approach against related work in literature; Sect. 3 provides an approach overview and describes the Smart City Ontology on which the approach relies; Sect. 4 describes the candidate indicators recommendation step, while in Sect. 5 we introduce the urban data exploration; Sect. 6 discusses the framework implementation; in Sect. 7 we present preliminary experiments; finally, Sect. 8 closes the paper.

2   Related Work

Our approach focuses on an ontology-based data exploration perspective for urban data, properly aggregated in the form of Smart City indicators, considering users' preferences and novel search interests and exploiting additional knowledge provided by the SCO. According to this perspective, our approach differs from *Ontology-Based Data Access* (OBDA), coping with the integration of heterogeneous data sources inside the Smart City [1,6,9,17], and from general purpose recommender systems as described in recent surveys (e.g., [11]), where the explorative viewpoint is not explicitly addressed. Furthermore, compared to approaches focused on Ontology-Based Data Warehouses (OBDW), which store analytical data, indicators, requirements and their semantics [14,15,23],

---

[1] http://www.bresciasmartliving.eu/.

our data exploration framework exploits indicators hierarchy and considers users' profiles to enrich the exploration experience, also considering the influence that the knowledge about indicators might have on users' activities. For what concerns the use of Semantic Web technologies in existing Smart City projects, ontologies have been used for *energy management*, where diagnostic models are defined to discover energy losses [25], or to perform optimisations for cost saving [2,5,12,21]; for *facility discovery*, to search for city facilities and services [4,12]; for *events monitoring and management* [3,19]. Existing approaches proposed recommendation of specific kinds of urban data, e.g., advising environmental recommendations to improve the quality of life of people [20], ontology-based exploration of crime data, that relies on Association Rule Mining [8], ontology-based visualisation of data or mobility [24]. The approach presented in [16] focuses on the ontology development phase. It defines a set of high level concepts, mapped to the ones of ontologies underneath twenty Smart City applications. In [18] an ontology, which models a Smart City as a composition of information objects, agents and measures, is proposed. In [23] a semantic characterisation of Smart City indicators is provided.

Differently from the aforementioned solutions, our proposal uses the concept of indicator to provide a holistic view over the entire Smart City domain, including a wider heterogeneous spectrum of urban data, such as data on energy consumption, environmental conditions, safety and security. Moreover, we foster personalised exploration of data for different categories of users (e.g., citizens, building managers, PA).

3   Approach Overview

The proposed approach is based on both ontology-based descriptions of Smart City indicators and users' profiles, as detailed in the following. According to the intrinsic modular nature of ontologies, the Smart City Ontology may be reused within other Smart City projects, after being properly extended to include the spectrum of concepts and relationships of the considered context. We do not discuss how to compute indicators from heterogeneous data sources, as this is out of the scope of the paper.

3.1   The Smart City Ontology

Figure 1 reports the main concepts and relationships of the Smart City Ontology[2] (SCO). The SCO provides a formal representation of Smart City indicators, with reference to the kinds of activities and users' categories for which indicators can provide relevant information. In particular, the SCO contains the definitions of the following main concepts and mutual relationships between them.

Indicators. Indicators represent an aggregation of urban data of interest for the citizens of the Smart City. Through the SCO, indicators are specified as

---

[2] The TBox of the ontology can be found at https://tinyurl.com/sco-onto (a free Web Protégé account is required).
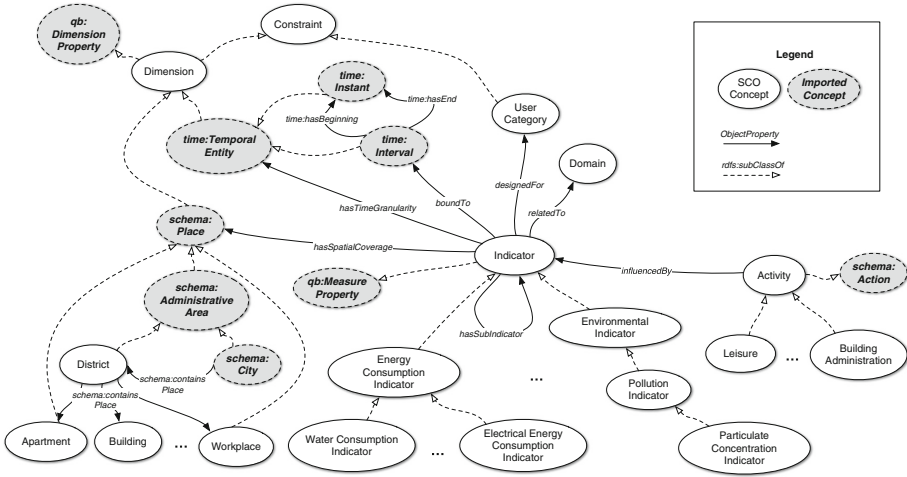
Fig. 1. A portion of the Smart City Ontology, containing main concepts and semantic relationships.

individuals of the `Indicator` concept or one of its sub-concepts in the indicators hierarchy. For instance, `ElectricalEnergyConsumptionIndicator` is defined as a sub-concept of `EnergyConsumptionIndicator` in Fig. 1. We denote with $\mathcal{I}$ the overall set of individuals of available indicators. In the SCO, an indicator is `relatedTo` a domain and is further specified through a set of `Constraint` individuals in the ontology.

Domains. A domain represents a concept used to limit the scope of an indicator, among the types of urban data that can be explored (e.g., environment, safety, energy, mobility). Therefore, an indicator $i \in \mathcal{I}$ is associated with a set of individuals of the `Domain` concept D_i (`relatedTo` relationship).

Constraints. In the SCO a constraint $c_i \in C_i$ can be either an user's category (e.g., citizen, building manager) or a dimension (i.e., time or space). On the one hand, an indicator is designed for specific users' categories (`designedFor` relationship). For instance, the indicator on electrical consumption of buildings is designed to be browsed and explored by the buildings manager. The user's category will enable personalised indicators filtering. On the other hand, an indicator can be `boundTo` a time interval (e.g., values of electrical consumption are available for the year 2017), may have a time granularity (`hasTimeGranularity` relationship) and may be defined at the city, street, district or more specific levels, such as buildings, workplaces and private apartments (`hasSpatialCoverage` relationship).

Activities. This concept is used to represent users' activities (e.g., leisure, building administration) that can be influenced by the knowledge provided by accessing an indicator. For instance, the `EnergyConsumptionIndicator` may provide useful insights for a building manager for implementing energy saving activities

in the administered buildings. Similarly, pollution indicators may prevent citizens from practising outside sport activities. To this aim, the `Activity` concept is connected in the SCO to the `Indicator` concept through the `influencedBy` property.

Other SCO Concepts. In defining the SCO, pivotal concepts from available foundation ontologies are exploited to: (i) represent geospatial concepts of the city (e.g., district, street) and users' activities (Schema.org[3] ontology), (ii) define temporal entities, used as analysis dimensions (Time[4] ontology) and (iii) characterise indicators as analytical data entities (Data Cube[5] ontology).

Formally, we can summarise an indicator i as a tuple $\langle ID_i, T_i, D_i, C_i \rangle (\forall i = 1, \dots, N_\mathcal{I})$, where ID_i is a unique identifier (i.e., an URI), T_i is the indicator type (e.g., `ElectricalEnergyConsumptionIndicator`), D_i is the set of domains individuals, C_i is the set of constraints, and where $N_\mathcal{I}$ is the number of indicators in $\mathcal{I}$.

3.2 Users' Profiles

Users are described according to a category (e.g., citizen, building manager), activities of interest, defined as individuals of the concept `Activity` or some of its sub-concepts, the types of indicators explored by the user through the interactions with the framework, defined through the concept `Indicator` or its sub-concepts. Different users may have access to different indicators: a citizen can select indicators concerning his/her apartment only, building managers can select indicators on their administered buildings only, energy managers can select indicators that only concern the workplaces they are responsible for, etc.

To this aim, after selecting the activities of interests (if any) during the registration to the framework, citizens, building managers and other categories of users have also to specify the places (e.g., an apartment, a building) they act in. Only urban data that has been aggregated within indicators associated with that places in the SCO (`hasSpatialCoverage` relationship) will be displayed. This has a two-fold advantage: (i) it enables data privacy preservation, for instance preventing building managers to visualise data on buildings they do not administer; (ii) it will be used to personalise indicators selection and data exploration, as explained in the next sections.

Formally, the profile $p(u)$ of a user $u \in \mathcal{U}$ can be summarised as $p(u) = \langle ID_u, cat_u, \mathcal{I}_u, A_u, P_u \rangle$, where ID_u is the identifier associated with the user's account, cat_u is the user's category, $\mathcal{I}_u$ is the set of individuals representing indicators that have been selected and explored by u in previous interactions with the framework, A_u is the set of activities of interest for the user, P_u is the list of individuals of concepts representing specific places where the user takes actions, namely individuals of `Building` for building managers, of `Apartment` for citizens, of `Workplace` for energy managers, etc. A registration wizard, starting

---

[3] http://schema.org/.
[4] http://www.w3.org/2006/time.
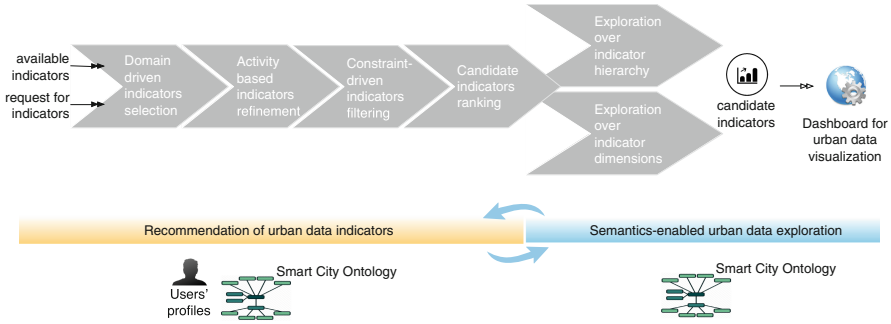[5] http://purl.org/linked-data/cube.

Fig. 2. The steps of semantics-enabled personalised data exploration.

from the user's category, prompts to the user proper masks to insert the instances defining the set P_u and associate them with districts. Such instances will be also inserted in the SCO to enable semantics-enabled urban data exploration, while only districts and upper level places (e.g., the city districts) are inserted in the ontology by domain experts. This reduces the complexity of ontology population and maintenance for platform administrators.

3.3   Urban Data Exploration Steps

The semantics-enabled approach proposed here for urban data exploration is articulated over two main steps, as shown in Fig. 2: (i) the overall set of available indicators is properly pruned, by taking into account both explicit requirements of the user as expressed in a search request and the user's profile, and then ranked (*recommendation of urban data indicators*); (ii) the list of selected indicators is presented to the user as starting point to enable a semantics-enabled personalised exploration of urban data (*semantics-enabled urban data exploration*).

Urban data exploration starts from a request formulated by the user u, denoted with $r(u)$, that contains the domains and indicators of interest, specified as individuals of `Domain` and `Indicator` concepts or sub-concepts of the SCO, respectively. Nevertheless, in order to provide support to the request formulation without demanding a detailed knowledge of ontology concepts and individuals, the framework allows the user to specify a set of keywords $K_r = \{k_{r1}, k_{r2}, \ldots, k_{rn}\}$. The set K_r is processed according to techniques aimed to match the keywords with ontology terms [22]. The adopted disambiguation procedure relies on WordNet[6] to retrieve synonyms, hypernyms and hyponyms of the keywords, and identifies a mapping between the input list of keywords and ontology individuals using probabilistic techniques. Following this approach, requests are processed in a more flexible way, to deal with the different levels of expertise (i.e., knowledge of the terminology and lexicon) users have. Beyond the domains and/or indicators explicitly indicated by the user in K_r, the user's

---

[6] https://wordnet.princeton.edu/.

profile is exploited to take into account other elements for indicators selection, namely the user's category cat_u, his/her activities of interest $\mathcal{A}_u$ and the set $\mathcal{I}_u$ of indicators explored in the past by the user.

Formally, we represent the request submitted by user u as follows: $r(u) = \langle D_r, \mathcal{I}_r, p(u) \rangle$, where D_r is the set of desired domains, $\mathcal{I}_r$ is the set of indicators of interest, $p(u)$ is the user's profile. D_r and $\mathcal{I}_r$ are the output of the WordNet-based disambiguation procedure.

Recommendation of urban data indicators is further organised in four sub-steps, namely:

– *domain-driven indicators selection*, in which a preliminary pruning of candidate indicators is performed based on the domains and indicators of interest specified in the request $r(u)$;
– *activity-based indicators refinement*, where the set of candidate indicators is further enriched considering the activities in the user's profile and the `influencedBy` ontological relationship;
– *constraint-driven indicators filtering*, where user's category and spatial constraints are exploited to further filter candidate indicators;
– *candidate indicators ranking*, performed by considering both the user's request and the history of indicators in the user's profile.

Urban data exploration, starting from the output of the recommendation of urban data indicators, is supported through the indicators hierarchy and semantic relationships of the SCO, enabling the users to browse the set of available indicators starting from the recommended ones, as explained in Sect. 5. Exploration actions may require to revise candidate indicators recommendation.

Finally, the user can visualise the actual values of one or more indicators in a numerical or graphical way by means of a suitable web-based dashboard. Even if we have implemented the web-based dashboard in the context of the BSL project, in the following the focus will be on the recommendation and exploration steps.

4 Recommendation of Urban Data Indicators

Candidate Indicators Selection. The selection process of candidate indicators is described by Algorithm 1. It takes as input the set $\mathcal{I}$ of all available indicators, the request $r(u)$ and the SCO. The output of the selection process is a set of candidate indicators, namely $\mathcal{I}_{cand}$, containing indicators that are compliant with the request.

Starting from the set $\mathcal{I}_r$ of indicators of interest as specified in the request $r(u)$ (see line 1), the selection process performs the domain-driven indicators selection (lines 2–6), the activity-based indicators refinement (lines 7–10), the constraint-driven indicators filtering (lines 11–18).

In the domain-driven indicators selection, the set D_i of individuals of the `Domain` concept, associated with each indicator $i \in \mathcal{I}$, is retrieved by considering

---

Algorithm 1. Candidate indicators selection

Input : Request $r(u) = \langle D_r, \mathcal{I}_r, p(u) \rangle$, set $\mathcal{I}$ of available indicators, the Smart
City Ontology SCO

Output: The set of candidate indicators $\mathcal{I}_{cand} \subseteq \mathcal{I}$

1 $\mathcal{I}_{cand} \leftarrow \mathcal{I}_r$;

2 **if** $(D_r \neq \emptyset)$ **then**

3 | **foreach** *indicator* $i \in \mathcal{I}$ **do**

4 | | $D_i \leftarrow \{domain \mid \texttt{relatedTo}(i, domain)\}$;

5 | | **if** $(D_i \cap D_r \neq \emptyset)$ **then**

6 | | | $\mathcal{I}_{cand} \leftarrow \mathcal{I}_{cand} \cup \{i\}$;

7 **if** $(A_u \neq \emptyset)$ **then**

8 | **foreach** *activity* $a \in A_u$ **do**

9 | | $\mathcal{I}_a \leftarrow \{indicator \mid \texttt{influencedBy}(a, indicator)\}$;

10 | | $\mathcal{I}_{cand} \leftarrow \mathcal{I}_{cand} \cup \mathcal{I}_a$;

11 $C \leftarrow P_u \cup cat_u$, both extracted from $p(u)$;

12 **if** $(C \neq \emptyset)$ **then**

13 | **foreach** $(c \in C)$ **do**

14 | | **switch** c **do**

15 | | | **case** $\texttt{UserCategory}(c)$

16 | | | | $\mathcal{I}_{cand} \leftarrow \mathcal{I}_{cand} \setminus \{i \mid i \in \mathcal{I}_{cand} \wedge \neg \texttt{designedFor}(i, c)\}$;

17 | | | **case** $\texttt{schema:Place}(c)$

18 | | | | $\mathcal{I}_{cand} \leftarrow \mathcal{I}_{cand} \setminus \{i \mid i \in \mathcal{I}_{cand} \wedge \neg \texttt{hasSpatialCoverage}(i, c)\}$;

19 **if** $(\mathcal{I}_{cand} == \emptyset)$ **then**

20 | $\mathcal{I}_{cand} \leftarrow \mathcal{I}_u$;

21 **return** $\mathcal{I}_{cand}$;

---

the `relatedTo` relationship in the SCO. If there is an overlapping between the sets D_i and D_r, then the indicator i is added to the set of candidates $\mathcal{I}_{cand}$.

In the activity-based indicators refinement, for each activity $a \in \mathcal{A}_u$ extracted from the user's profile, the `influencedBy` relationship in the SCO is used to retrieve additional candidate indicators.

During the constraint-driven indicators filtering, each indicator $i \in \mathcal{I}_{cand}$ is analysed to filter out candidate indicators that are not compliant with the user's category cat_u and are not available for the places P_u where the user takes actions, which are both extracted from the user's profile as a set of constraints C (line 11). Considering the `designedFor` relationship in the SCO to get the individuals of concept `UserCategory` that are semantically related to i, cat_u is used to filter out non relevant indicators from $\mathcal{I}_{cand}$ (lines 15–16). Similarly, indicators that are not available for the places P_u are discarded considering the `hasSpatialCoverage` relationship (lines 17–18). If the set of candidate indicators is empty, then it is populated with indicators included in $p(u)$ (lines 19–20). The

rationale behind this choice is that the set $\mathcal{I}_u$ traces past exploration history of the user providing known, albeit not novel, candidate indicators.

Candidate Indicators Ranking. Once the set $\mathcal{I}_{cand}$ has been identified, candidate indicators are properly ranked by combining two different criteria: (a) the similarity with the user's request $r(u)$ (denoted with Sim_{req}) and (b) the similarity with indicators selected by the user in past exploration activities (denoted with Sim_{past}). The rationale is that the latter criterion is used to maintain a certain compliance with usual interests of the user, as represented through the set $\mathcal{I}_u$ of indicators in his/her profile. On the other hand, only relying on past choices of the user would penalise user's new interests. This can be viewed as a variation of the cold start problem [11] that affects recommendation systems and is balanced by the first criterion.

Specifically, for what concerns (a), the similarity measure is calculated by estimating the overlap between the set D_r of domains, included in $r(u)$, and the domain(s) associated with each candidate indicator $i_{cand} \in \mathcal{I}_{cand}$. The second kind of similarity, that is based on the proximity of each i_{cand} with respect to indicators in $\mathcal{I}_u$, is computed by measuring the similarity between places and domains that are shared by the computed indicators.

The ranking function $\rho : \mathcal{I}_{cand} \longmapsto [0,1]$ is computed as follows:

$$\rho(i_{cand}) = \alpha \cdot Sim_{req}(D_r, D_{i_{cand}}) + \beta \cdot Sim_{past}(\mathcal{I}_u, i_{cand}) \tag{1}$$

where $\alpha + \beta = 1$ (in our preliminary experiments we set $\alpha = \beta = \frac{1}{2}$), and $D_{i_{cand}}$ is the set of domains associated with $i_{cand} \in \mathcal{I}_{cand}$.

The term Sim_{req} is computed by applying the Dice coefficient, which estimates the overlapping between two sets, as follows:

$$Sim_{req}(D_r, D_{i_{cand}}) = 2 \cdot \frac{|D_r \cap D_{i_{cand}}|}{|D_r| + |D_{i_{cand}}|} \tag{2}$$

Regarding Sim_{past}, this similarity is evaluated as:

$$Sim_{past}(\mathcal{I}_u, i_{cand}) = MAX_{i_u \in \mathcal{I}_u} Sim(i_u, i_{cand}) \tag{3}$$

where $Sim()$ combines the similarity coefficients, equally weighted, between the two indicators according to their domains and their places:

$$Sim(i_u, i_{cand}) = \frac{|P_{i_u} \cap P_{i_{cand}}|}{|P_{i_u}| + |P_{i_{cand}}|} + \frac{|D_{i_u} \cap D_{i_{cand}}|}{|D_{i_u}| + |D_{i_{cand}}|} \tag{4}$$

where P_{i_u} (resp., D_{i_u}) is the set of places (resp., domains) associated with the indicator i_u and $P_{i_{cand}}$ (resp., $D_{i_{cand}}$) is the set of places (resp., domains) associated with the indicator i_{cand}. Please note that all these similarity coefficients are in the range $[0,1]$.

5   Semantics-Enabled Personalised Urban Data Exploration

The SCO is used to support both the candidate indicators selection and ranking, as discussed above, and the exploration of indicators. The latter one is meant as

an iterative process guided by the SCO organisation of concepts and relationships, and including exploration of both semantic description of indicators and of actual values accessible through the web-based dashboard.
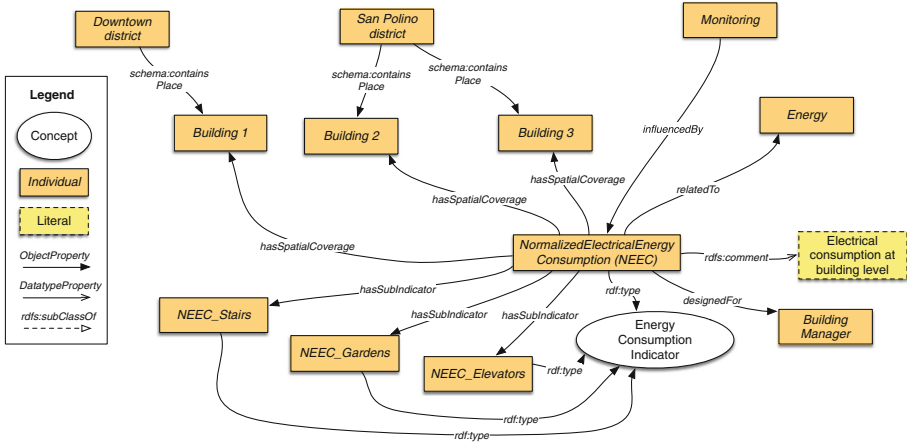

Fig. 3. Example of candidate indicator and related properties.

Starting from candidate indicators returned through the selection step, the users of the platform can further explore other indicators being guided by the semantic relationships in the SCO. Exploration can be performed according to different perspectives, given the knowledge structure in the ontology: (a) exploration over the indicators hierarchy; (b) personalised exploration over the indicators dimensions. Let us explain how this can be done in our framework by a simple demonstration scenario.

Let's consider the user John in the motivating example, who is the manager of three buildings (namely `Building 1`, `Building 2` and `Building 3`) located in two districts of the city. In particular, `Building 1` is located in the city downtown, while `Building 2` and `Building 3` are located in the modern district of San Polino. Since John is interested in monitoring buildings, during the registration to the platform he specifies the activity `Monitoring` in his profile. Moreover, he specifies what are the administered buildings and associates them with the districts they are located in, as part of his profile. Buildings are also inserted into the SCO and linked to the districts by means of the `schema:containsPlace` relationship. In order to have an insight on the status of the buildings he administers, for instance to evaluate whether replacing standard lamps with less energy-demanding LED ones, John logs in to the platform and asks for consumption indicators, specifying the keywords $K_r = \{$energy, consumption$\}$. The platform processes the request as explained in the previous section and returns, among the others, the indicator `NormalizedElectricalEnergyConsumption` (NEEC), which reports electrical consumption normalised with the number of apartments

in the building. The indicator is selected because it is both compatible with the keywords given in the request and associated with the activity `Monitoring` in the ontology. Semantic description of the indicator `NEEC` is shown in Fig. 3.

If John decides to explore the indicators hierarchy, he may select the `NEEC` indicator and the framework suggests him more specific indicators `NEEC_Stairs`, `NEEC_Elevators` and `NEEC_Gardens`, which are related to the `NEEC` indicator through the `hasSubIndicator` relationship in the ontology. Since John's focus is on evaluating the electrical consumption of the lighting plants of stairs, he selects `NEEC_Stairs`.

Personalised exploration over the indicators dimensions exploits the semantic relationship `schema:containsPlace` that relates each others individuals of `schema:Place` concept or its sub-concepts. In particular, knowledge on the spatial coverage of indicators is obtained through the `hasSpatialCoverage` relationship. Starting from indicators previously selected for the John's building, either `NEEC` or one of its sub-indicators, the containment relationship that relates John's buildings with districts is exploited. Therefore, John could choose to visualise the average consumption provided by the indicators for the buildings of the districts, in order to compare his buildings against other ones having similar characteristics or using different lighting solutions. Similarly, indicators for John's buildings could also be suggested over several years (`boundTo` relationship) or over different time granularities (e.g., years, months, days), according to the `hasTimeGranularity` relationship. Comparison between indicators may stimulate John to consider the replacement of energy consuming light bulbs with modern LED lamps in shared spaces, after analysing the affordability of the expenses, with respect to the ones sustained by other similar buildings.

6   Three-Layered Framework Architecture

Figure 4 shows an overview of the semantics-enabled data exploration framework architecture. The framework is developed with web-based technologies and is organised over multiple layers. It has been preliminarily applied in the context of the Brescia Smart Living (BSL) Italian project. Data on field, collected from domain-specific platforms through IoT technologies, as well as data from sources external to the BSL project (weather data, pollution data, etc.) are loaded into the BSL platform database (*BSL Platform Layer*). Data is transferred on the BSL platform using RESTful services, SOAP-based services and MQTT Agents. The *Semantic Layer* enables personalised urban data exploration, as described in Sects. 4 and 5. The *User Access Layer* includes a web-based Smart City Dashboard to be used by citizens, PA and other users to explore urban data.

Using the web browser, users can register themselves and update their profile. Within each layer of the platform, proper Web Services, implemented in Java and deployed under the Apache TomEE[7] application server, elaborate the incoming requests from users to retrieve the set of recommended indicators. The Smart
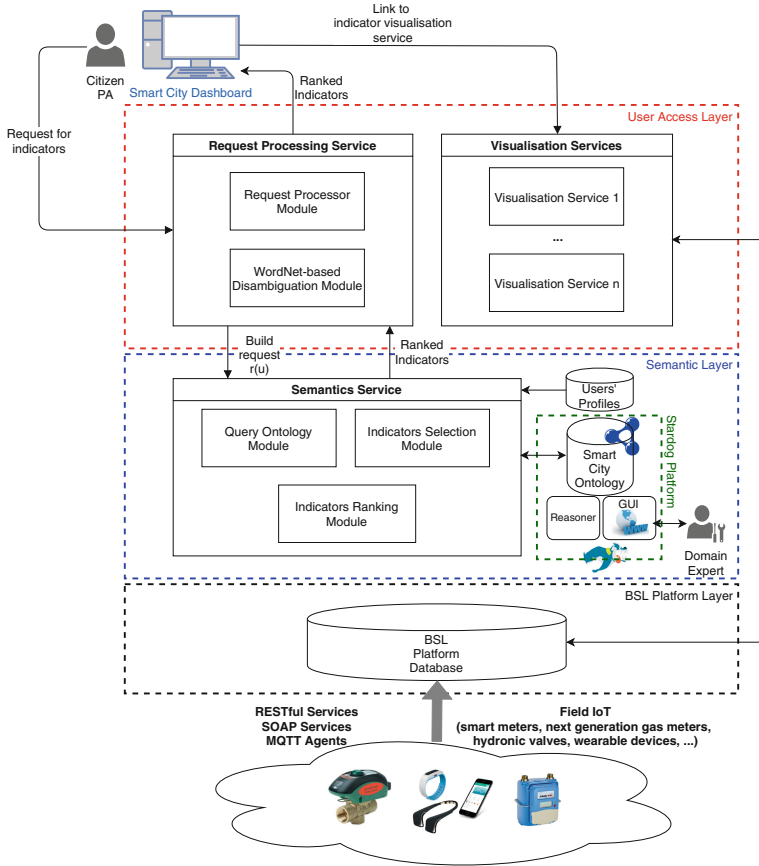
---

[7] http://tomee.apache.org/.

Fig. 4. Web-based architecture of the semantics-enabled data exploration framework.

City Ontology is deployed in OWL using Stardog[8], a NoSQL graph database based on W3C Semantic Web standards. The Stardog Platform supports domain experts in order to maintain the ontology (concepts, relations and individuals, including the insertion of new indicators individuals), interacting with the web-based administration console provided by the platform.

In this section, we introduce the main Web Services, located within the layers of the framework, invoked to process users' requests for indicators recommendation and exploration. Here, we present only the services and their composing modules at a high level, without lingering much on technical details, showing the interaction flow triggered by the user (either a citizen or PA) when issuing a request to the framework. The sequence diagram of Fig. 5 illustrates the interactions (i.e., invocation of methods and exchanged parameters) between the main

---

[8] https://www.stardog.com/.

modules of the framework in order to perform the indicators recommendation step. These modules are implemented as services described in the following.
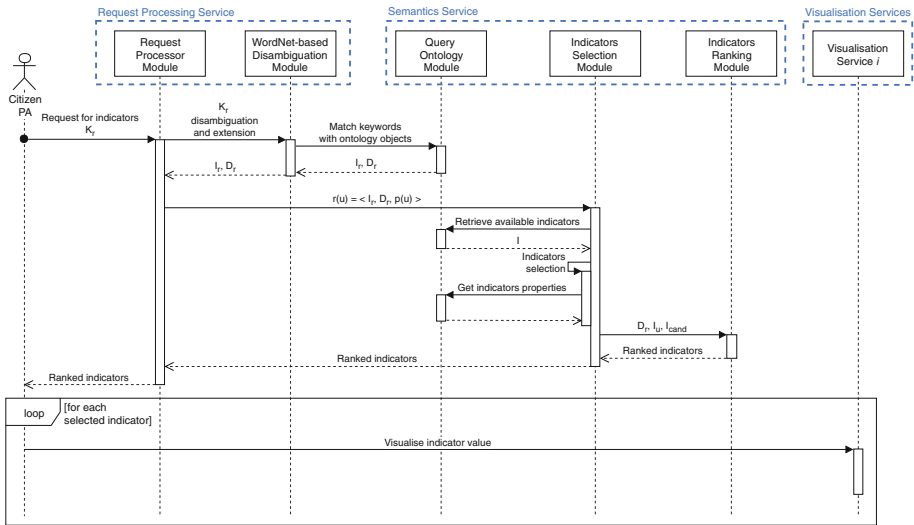

Fig. 5. Interactions between the main modules of the architecture.

Request Processing Service. This service is in charge of processing the incoming users' requests, starting from the set of keywords K_r and the user's profile $p(u)$. It is composed of the following modules: (i) the *Request Processor Module*, that routes the requests to the other services of the platform, handles partial results and returns the set of recommended indicators to the user; (ii) the *WordNet-based Disambiguation Module*, that looks up WordNet database to accomplish the semantic enrichment and disambiguation of the keywords contained in the query.

Semantics Service. The modules embedded in this service operate on the Smart City Ontology, to retrieve and select the candidate indicators, compliant with the user's request. Specifically: (i) the *Query Ontology Module* queries the ontology through the API provided by the Stardog Platform; (ii) the *Indicators Selection Module*, performs the selection of candidate indicators; (iii) the *Indicators Ranking Module* implements the techniques exposed in Sect. 4 to rank the set of candidate indicators. The Query Ontology Module, embedded into the Semantics Service, is in charge of issuing SPARQL queries over the SCO; specifically, it handles two different types of queries: (i) *selection* queries and (ii) *boolean* queries. The former ones (also known as *SELECT* queries) are used to retrieve individuals from the ontology (e.g., to collect the set of indicators apt to citizens) whereas the latter ones (also known as *ASK* queries) are useful to query the ontology to get true/false answers (e.g., to check whether the NEEC indicator

is available for citizens). SPARQL queries can benefit from the underlying reasoning engine, provided by the Stardog Platform, to enrich their results when computing hierarchies (e.g., to build the lineage of an indicator, up to its topmost ancestor) or to infer additional background knowledge.

Visualisation Services. This group of services comprises legacy modules that lie in the visualisation engine of the platform. Each indicator has a link which triggers a proper Web Service inside the platform, that enables the visualisation of the indicator values onto the Smart City Dashboard, retrieving the data from the BSL platform database.

7   Preliminary Experiments

Preliminary experiments on the proposed framework aim at demonstrating its effectiveness in supporting candidate indicators selection for a given request r.

To this aim, we compared our framework against a baseline keyword-based approach, where keywords have been properly expanded with synonyms using WordNet lexical system. In particular, given a request r, precision $P(r)$ (i.e., the number of relevant indicators compared to the total number of returned indicators, in the context of a search operation) and recall $R(r)$ (i.e., the number of relevant indicators returned among the search results, compared to the total number of relevant indicators) are measured for quantifying the effectiveness of the candidate indicators selection. To measure precision and recall, we used a SCO composed of 57 concepts (and, among them, 30 indicators), 104 individuals, 207 object and datatype properties. Table 1(a) reports average precision and recall values of our approach compared to a keyword-based search. Precision and recall values have been computed on two kinds of requests: (a) requests where the user specified a set of keywords to identify desired domains and indicators, and the user's profile does not contain any activity or preferential indicator (r_A); (b) requests where the user presents a richer profile (containing category, activities and preferential indicators), but specifies a few keywords in the keyword set K_r, that only correspond to individuals of the Domain concept (r_B). Five requests for each type have been issued and average values have been computed.

The second kind of request is used to demonstrate how relationships within the SCO are effective in improving precision and recall for indicators selection. In fact, with respect to the keyword-based approach, our framework enables a better precision by refining the set of candidate indicators based on the user's category, the specified domain(s) and other ontological relationships. On the other hand, recall is increased by exploiting the relationships between other elements of the user's profile (i.e., activities and preferential indicators) and the available indicators in the ontology, thus including among search results candidate indicators that are not described with the keywords specified in the user's request or with keyword synonyms as extracted from WordNet. Since both the compared approaches use WordNet to perform keywords disambiguation and the same keywords across the approaches have been used during tests, difference

Table 1. (a) Average precision and recall values obtained for the preliminary evaluation; (b) average execution time (in msec) for indicators recommendation.

	Keyword-based search	Ontology-based search
$P(r_A)$	0.49	0.99
$R(r_A)$	0.97	0.98
$P(r_B)$	0.33	0.94
$R(r_B)$	0.27	0.93

(a)

Request type	Indicators selection	Keywords disambiguation	Ranking	Total
r_A	1454	1105	805	3361
r_B	269	1056	880	2205

(b)

in average precision and recall is due to the knowledge structure provided by the ontology.

The formulation of the request as a set of keywords, instead of asking the user to specify required properties and constraints, enables more flexibility, since it does not demand for a detailed knowledge of the ontology, its concepts and relationships. Furthermore, the processing time required to expand keyword sets with the use of WordNet is affordable and acceptable for the considered exploration scenarios. Table 1(b) contains average execution time (in msec) for indicators recommendation (including time spent for the WordNet-based disambiguation of keywords and for ranking) in the preliminary evaluation. Tests have been performed on a Windows-based machine equipped with an Intel i7 2.00 GHz CPU, 8 GB RAM, SSD storage.

Usability tests are being performed to check the capability of the framework in facilitating user's access to urban data through the suggestion of candidate indicators. To perform usability tests, we considered a population of users using metrics such as the number of exploration steps needed to obtain desired data, number of fails, number of successful explorations. Usability experiments are being carried on within the Brescia Smart Living project until September 2018. Currently, the framework is being tested, with satisfaction, by a sample of users in two districts, a modern one (San Polino), where new generation smart meters have been installed, and a district in city downtown, more densely populated and presenting older buildings. The framework will be also used by other partners involved in the project as representatives of PA (in particular, the Municipality of Brescia, Italy), utility and energy providers.

8   Conclusions and Future Works

In this paper, we described a framework for enabling personalised exploration of urban data. The framework relies on the knowledge structure provided by a

Smart City Ontology (SCO) and the information contained within the citizens' profiles. The SCO is used to properly represent indicators in terms of concepts, hierarchies and semantic relationships, thus they can be used to facilitate exploration by exploiting the knowledge structure. Moreover, the concept hierarchies in the ontology allow exploration at different granularity levels. Candidate indicators are recommended if compliant with user's profile, ensuring a personalised selection and exploration over the set of available Smart City indicators. The semantics-enabled personalised urban data exploration is articulated over two steps: (i) indicators are properly recommended taking into account both explicit requirements of the user as expressed in a search request and the user's profile; (ii) recommended indicators are exploited as starting point to set up interactive exploration of urban data. Future efforts will be devoted to perform further experiments (e.g., comparing [22] with other matching techniques) and to extend the set of semantic relationships in the SCO as follows: (a) further relationships between indicators will be identified (e.g., to assert that two or more environmental indicators must be jointly monitored due to their harmful impact on the ecosystem); (b) strategies to promote the users' virtuous behaviours will be studied and implemented on top of the relationships, providing advices for healthy activities that should be practised by users. This will be accomplished by collecting and formalising additional knowledge about users' lifestyle, and then enriching the SCO with specific background semantics. Finally, support to the insertion of indicators individuals, that is currently performed by domain experts given their skill in the application context, will be further developed as well.

References

1. The GrowSmarter project. http://www.grow-smarter.eu/home/
2. OPTIMising the energy USe in cities with smart decision support systems. http://optimus-smartcity.eu/
3. The Res Novae project. http://resnovae-unical.eu
4. The San Francisco Park project. http://sfpark.org
5. The BESOS project: Building Energy decision Support systems fOr Smart cities. http://besos-project.eu
6. The ROMA project: Resilience enhancement Of a Metropolitan Area. http://www.progetto-roma.org
7. Anttiroiko, A.V.: City-as-a-platform: the rise of participatory innovation platforms in finnish cities. Sustainability **8**(9), 922 (2016)
8. Balasubramani, B.S., Shivaprabhu, V.R., Krishnamurthy, S., Cruz, I.F., Malik, T.: Ontology-based urban data exploration. In: Proceedings of the 2nd ACM SIGSPATIAL Workshop on Smart Cities and Urban Analytics (UrbanGIS), pp. 10:1–10:8 (2016)
9. Bellini, P., Benigni, M., Billero, R., Nesi, P., Rauch, N.: Km4City ontology building vs data harvesting and cleaning for smart-city services. J. Vis. Lang. Comput. **25**(6), 827–839 (2014)
10. Bianchini, D., De Antonellis, V., Garda, M., Melchiori, M.: Semantics-enabled personalised urban data exploration. In: Proceedings of 19th International Conference on Web Information System Engineering (WISE) (2018). Accepted for publication

11. Bobadilla, J., Ortega, F., Hernando, A., Gutirrez, A.: Recommender systems survey. Knowl.-Based Syst. **46**, 109–132 (2013)
12. Brizzi, P., Bonino, D., Musetti, A., Krylovskiy, A., Patti, E., Axling, M.: Towards an ontology driven approach for systems interoperability and energy management in the smart city. In: International Conference on Computer and Energy Science (SpliTech), pp. 1–7 (2016)
13. Chauhan, S., Agarwal, N., Kar, A.: Addressing big data challenges in smart cities: a systematic literature review. Info **18**(4), 73–90 (2016)
14. Fox, M.S.: PolisGnosis project: representing and analysing city indicators. In: Enterprise Integration Laboratory, University of Toronto Working paper (2015)
15. ISO: Sustainable development of communities - Indicators for city services and quality of life. Standard, International Organization for Standardization (2014)
16. Komninos, N., Bratsas, C., Kakderi, C., Tsarchopoulos, P.: Smart city ontologies: improving the effectiveness of smart city applications. J. Smart Cities **1**(1), 1–16 (2015)
17. Lopez, V., Stephenson, M., Kotoulas, S., Tommasi, P.: Data access linking and integration with DALI: building a safety net for an ocean of city data. In: Arenas, M., et al. (eds.) ISWC 2015. LNCS, vol. 9367, pp. 186–202. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-25010-6_11
18. Psyllidis, A.: Ontology-based data integration from heterogeneous urban systems: a knowledge representation framework for smart cities. In: Proceedings of the 14th International Conference on Computers in Urban Planning and Urban Management (2015)
19. Rani, M., Alekh, S., Bhardwaj, A., Gupta, A., Vyas, O.P.: Ontology-based classification and analysis of non-emergency smart-city events. In: 2016 International Conference on Computational Techniques in Information and Communication Technologies (ICCTICT), pp. 509–514 (2016)
20. Riga, M., Kontopoulos, E., Karatzas, K., Vrochidis, S., Kompatsiaris, I.: An ontology-based decision support framework for personalized quality of life recommendations. In: Dargam, F., Delias, P., Linden, I., Mareschal, B. (eds.) ICDSST 2018. LNBIP, vol. 313, pp. 38–51. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-90315-6_4
21. Rossello-Busquet, A., Brewka, L.J., Soler, J., Dittmann, L.: OWL ontologies and SWRL rules applied to energy management. In: Proceedings of the 2011 UkSim 13th International Conference on Computer Modelling and Simulation, pp. 446–450 (2011)
22. Royo, J.A., Mena, E., Bernad, J., Illarramendi, A.: Searching the web: from keywords to semantic queries. In: Third International Conference on Information Technology and Applications (ICITA), pp. 244–249 (2005)
23. Santos, H., Dantas, V., Furtado, V., Pinheiro, P., McGuinness, D.L.: From data to city indicators: a knowledge graph for supporting automatic generation of dashboards. In: Blomqvist, E., Maynard, D., Gangemi, A., Hoekstra, R., Hitzler, P., Hartig, O. (eds.) ESWC 2017. LNCS, vol. 10250, pp. 94–108. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-58451-5_7
24. Sobral, T., Galvão, T., Borges, J.: Semantic integration of urban mobility data for supporting visualization. Transp. Res. Procedia **24**, 180–188 (2017)
25. Tomašević, N.M., Batić, M.C., Blanes, L.M., Keane, M.M., Vraneš, S.: Ontology-based facility data model for energy management. Adv. Eng. Inform. **29**(4), 971–984 (2015)