Databases and Information Systems VIII
H.-M. Haav et al. (Eds.)

doi:10.3233/978-1-61499-458-9-303

Urban Data Querying: The City of Petrozavodsk Waste Problem

Olga IANIUK [a,1], Boris NOVIKOV [a]
[a] *Mathematics and Mechanics Department, St. Petersburg State University, Russia*

Abstract. Urban data mining is a promising tool for understanding complicated city systems. Indeed, data in urban spaces acquired from sensors and the web provides insights which, when used correctly, help city operation and development. Hopefully, it must improve the quality of citizens' life drastically in the present and future. However, data specifics involve particular difficulties in data management which results in its limited usage among specialists. Development and enhancement of techniques and tools that is used for data operation while tackling real urban challenges allow one both gain wider acceptance of such methods and increase impact of decisions made.

In this paper, an approach treating a certain declarative query language for management data on municipal solid waste is introduced. The information system high-level design issues implied by the urban data querying paradigm are proposed.

A case study is conducted in Petrozavodsk, Republic of Karelia, Russia, aiming to make the process of introduction of waste separate collection more efficient. The urban data management approach is supported by a real-life task of OpenStreetMap objects of the Petrozavodsk area manipulation which is essential for more complicated urban analysis.

Keywords. urban data management, analytical system, declarative query language, waste management system, Petrozavodsk

Introduction

The notion of "smart city" becomes a favorite one for an increasing number of municipal authorities and is often associated with the city's "mission control center", however this outlook seems to be too parochial.

There has not been any *totally smart city* for now, and there is no certainty that it will ever appear. Instead of elaborating *not enough smart cities*, urban scientists and practitioners could detect urban development tasks for any particular region using urban informatics or urban computing methodology.

Big urban data mining as an urban informatics technique allows one to put into practice a more comprehensive approach to understanding city functions and problems than ever. It has considerable importance for urban practitioners, policy makers and citizens as it is they who can become smarter while making data driven decisions.

[1] Corresponding Author: Olga Ianiuk, Saint Petersburg State University, 28 Uinversitetsky Prospect, 198504, St. Petersburg, Russia; E-mail: olya.ianiuk@gmail.com.

Moreover, the notion of "big urban data" should also be considered more widely than the common "big data" buzzword. It is a domain driven context that accounts for: 1) significant variation of data sources in terms of heterogeneity, data models and availability; 2) tagging of data streams to real time and space; 3) importance of privacy and visualization issues; 4) invariable growth of data volume. Obviously, all of these items and not only the last one, add prefix "big" to "urban data".

Since information and communication technologies infrastructure expands, more and more agents of urban systems have a potential to derive data and it becomes more complicated to address information extraction and integration problems as preliminary stages of data mining. In addition, there is a shortage of appropriate tools assisting efficiently to handle big urban data.

On the other hand, if intrinsic city's problem-solving practice involves compulsory operating with high volume and complex data, it should both enforce urban data management techniques and improve the quality of decisions made.

Municipal solid waste (MSW) management inefficiency in Petrozavodsk, the urban center of Karelia region, is a real example of a serious urban problem. The main concern is introduction of separate MSW collection and its recovery which runs at a moderate pace for now.

Our research shows that asynchronous actions of different city groups is one of the reasons of negligible enhancement of MSW management system (MSWMS). Providing more valuable information on Petrozavodsk urban space and communities and their behavior, needs, aspirations, strengths and opportunities within the problem framework for stakeholders and citizens as well can make the introduction process more efficient. The present paper contains results of a preliminary stage of ongoing research project on an information system based on a declarative query language which addresses this issue. It is an expanded and improved version of a paper by the same authors [7] that comprises new tables, figures and descriptions.

The rest of the paper is structured as follows. In Section 1, the urban problem is introduced and in Section 2, urban data specifics are delineated. Sections 3 and 4 are devoted to analysis framework description. In Section 5, a practical example of possible analysis is included. Finally, the first research project results are discussed in Section 6, related work is reviewed in Section 7.

1. Petrozavodsk. Waste Management System Problem

Experts and researchers consider an introduction of MSW separate collection and recovery as an obvious need for most of Russian cities [1,22,19]. The common solid waste treatment practice in the country is disposal by landfilling, which generates about 200 million metric tons of excessive waste each year. It results in negative environmental impact and misuse of resources and energy [8,21].

Additionally, the number of cities with population over 200 000 and high purchasing power[2] in the North-West of Russia is significant (see Table 1). The volume of MSW they can produce is substantial and affects both Russia and neighbor countries of Baltic and Barents Regions [19,6].

[2]From Top-100 Russian cities by "Kommersant" [http://bit.ly/1mu1Zu8]

Table 1. Cities Ranking.

City	Population	Purchasing Power
Saint Petersburg	4880000	> 0.7
Kaliningrad	433000	0.7
Murmansk	305000	0.7
Petrozavodsk	260000	0.5

Since Russia had ratified the Kyoto Protocol, some measures for MSWMS improvement were adopted by government agencies at municipal level including Saint Petersburg [19]. Unfortunately, most efforts failed due to insufficient preparation and poor management [1].

However Petrozavodsk is marked by more comprehensive approach dealing with the problem.

The city has a population of 260 000 and stretches along the western shore of Lake Onega, the second largest in Europe after Ladoga. The local landfill site[3] occupies 36 hectares and is located in 16 km from the city center, 6 km from Lake Onega and about a kilometer from the Neyloksa river that flows into Lake Onega (see Figure 1).

Recently Petrozavodsk has faced decreasing capacity of the dump which causes concerns of regional authorities as well as a local waste management company Autospetstrans [10,14]. Besides upgrading the final disposal of MSW at the landfill [14], an international W.A.S.T.E. [10] project was established. It includes educational initiatives in elementary schools and kindergartens and vocational rehabilitation of waste management specialists and school teachers [23,13].

Figure 1. Petrozavodsk: Local Landfill Site

A waste processing plant necessary for Petrozavodsk and other cities in Karelia [17] has been planned from 2004 [13].

Meanwhile, the demand for secondary materials (paper, cardboard, plastics and construction waste) increases and is not completely satisfied. Ekolint company installs MSW receptacles and transports them free of charge within the framework of W.A.S.T.E. project [23]. Citizens are quite interested in MSW sorting; there is a number of online communities that coordinate households' independent MSW separate collection initiatives. Even though residents tend to be enthusiastic, the enterprises endure unjustified

[3] Actually the term "dump" is more appropriate as the site is open to precipitation [http://bit.ly/1e8dXQ0]

costs because of disjointed activities. Furthermore, public awareness and municipal support are still inadequate.

There are two more issues related to the industry improvement: imperfect regulatory framework [22,1,9] and unavailable/inaccurate economic, social, environmental and waste data for decisions on MSW treatment refinement [9,19].

Under such circumstances, one of the first steps to modify MSWMS is to link available data on the Petrozavodsk urban area components in order to make further decisions to be driven by insights potentially obtained.

2. Data Specification Framework

One of the key constraints to apply urban data management approaches in Russian cities is that accessibility and availability of datasets produced automatically and routinely from sensors or other suitable sources built-in space are exceptions rather than a rule. Thus, a major challenge for the MSW problem in Petrozavodsk is, but not limited to, a complete cycle of feature engineering from determination of data features and corresponding sources to data aggregation and pure statistics acquisition. Therefore, a flexible workable environment should be built to satisfy analysts' tasks arises.

Arranging the environment, the following specifics of urban data might be also taken into consideration. First, most of data comprehends temporal and/or spatial stamps as they are inherent measurements of urban space [3]. Therefore management of information feeds from environmental sensors, social networks or other services as well as certain aspects of time series analysis should be explicated during any relevant research.

In case of development of information systems for Petrozavodsk, one may predict MSWMS enhancement in the area, at least due to a possible opportunity to real-time analysis of MSWMS components. Simultaneously, approximate real-timeness is the next essential point of urban data.

If analytical or decision-making information system is built, an opportunity to easily visualize data, for example during exploratory analysis is valuable. The latter, in combination with high-relevance information mode, enforces additional requirements on data management tools and information systems incorporated them.

Another significant consideration relates to urban space users' data security and privacy support in time of research or results publishing [2].

Despite the deficiency of data sources mentioned beyond, it does not entirely mean any deficiency of data volume likely operated. Approaches to master problems associated with urban data specifics should be able to scale in order to handle growing data volume, for instance during longitudinal studies.

Some examples of information sources that are relevant for the Petrozavodsk area analysis are shown in Table 2.

To account for these limitations, appropriate tools for acquiring, integration and analysis should be applied. There is a number of commercial and open source solutions (i.e., ArcGIS, QGIS) which allow one to manipulate with relevant urban data. Essentially, options they proffer for analysts are mostly predefined which is not totally sufficient for flexible formulations of information needs over various data sources.

In agreement with flexibility requirements, single applicable modules of already existing tools as well as self-made scripts ensure independence for urban space investiga-

Table 2. Relevant Information Sources.

Source Name	Potential Usage
OpenStreetMap	physical objects representation
Microblogging: Twitter, Instagram, 4square etc.	sentiment analysis in the city context
Social networks: Vkontakte, Facebook etc.	information spread in the city context
...	...

tors. Meanwhile, the complete development of purposeful software is a kind of desirable limit for this method, it assumes intensive use of resources (i.e., financial, human, technological, time) and appears to be comprehensive to reproduce [4,20].

During the current research project, the analytical information system based on a declarative query language introduced in [12] is built.

It might arrange both an environment for expressing numberless information needs and germane automation of big data management tasks. Algebra operations established by the authors as a basis for queries is an intermediate language that is able to be optimized while querying heterogeneous data. However, these academic developments require extendibility by being adapted to the local context and producing more user-friendly experience.

3. Information System Design Issues

Elaborating event-driven three-tier architecture to handle large-scale data and traffic is an ordinary practice. Meanwhile it should be well-designed in terms of risks assessment, bottlenecks finding, infrastructure choice etc., even for widespread business.

With a peculiarity of urban data elucidated in Section 2 and recent emergence of the field, urban analytical system design becomes reasonably complex. However, high-level declarative query scripting allows one to design complicated processing and perform it easily by means of elaborated operations and analytical techniques of a querying engines implemented and incorporate one's own modules to meet domain specific needs [5,25]. In the present work, the instruments used as a basis (see Section 4) imply particular constructive properties of the building analytical system.

Unlike in the three-tier architecture, the instruments allow one to relax a requirement of application processing and data management functions being logically separated, for instance to reduce a tier. It is a hierarchical (nested) structure of queries that comprises both analytics and OLAP-cubes operation.

Besides, the declarative querying makes this process rather clear which is in correspondence with analysts' goals and opportunities.

In accordance with the data framework delineated in Section 2, subsets of integrated data needed for analytical tasks satisfaction are not defined beforehand. Thus, Extract-Transform-Loading processes (ETL) logical rules formalization becomes quite sophisticated. However, the algebra utilizing affords high-level "on-the-fly" ETL, since data sources are defined and extraction operations are performed at almost the same time with targeting query issued.

Whether business rules and data are kept in one container or not, data logical partitioning is a must. Furthermore, the processes of partitioning, removing or backups might be easily implemented with the help of the instruments through high-level scripting. That is, subsets of data enveloped by q-sets with objects of intended scores throughout different phases might be extracted by simple filters or interchangeable operations as inputs for a specified system function. In Listing 3, an example of prior removing instructions is shown.

In addition, logical data management depends on primary objects structure, thus source-case-like scenarios of data representation keep the building system at workable level.

In case of the system scaling necessity, the importance of which for big data processing is illustrated in [25], distributed nodes might be horizontally added by requisite declarative operations. Hence, the system design considerations have much in common with in-memory computing platform concepts recently established, having a great potential to address large-scale data handling effectively.

By these means, the implements in [12,24] exerting to the certain problem-solving urban task provide bottom-up large-scale data processing.

4. Declarative Specification Framework

It is important to highlight some aspects of the concept introduced in [12] in order to expound properties of the building analytical system.

A q-set as a central abstraction which is defined as a triplet $\langle q, B, S \rangle$, where q is an issued query, B is a base set of evaluated objects and S is a set which contains scores of elements in B corresponding to q.

A q-set is a notion that presents a model for similarity matching, namely encapsulates both a query and the result of its evaluation on objects in terms of similarity as a proximity measure in a certain space. The similarity is represented by a scoring function and depends on particular algebraic operations definition used in queries composition. Scores of objects might be also considered as values of internal/external functions integrated.

A q-set concerns the concept of fuzzy sets and scores might be interpreted as a membership function values being forced into the interval $[0,1]$. External scores need a calibration procedure.

Generally, the operations require no assumptions on their representation, although only q-sets must be treated as input/output arguments. Thus, a nested structure of query is derived: any final q-set subsumes previous q-sets as the base set along with primary q-sets related to raw data by queries of extraction. Then a final set is available as input either for external functions or the first operation in a new query. Also they might be stored at a q-sets storage, likewise any other intermediate or final resulting q-sets [11].

Obviously, primary objects (those obtained from information sources) must have cross-source identity and certain attributes to be evaluated by extraction queries. Indeed, the language of objects retrieved from primary data sources is not defined, meanwhile XML-like queries are formalized in the engine implemented [11].

There are several operations that should be underscored for use in explaining example of Section 5.

The most common operations are different kinds of filters. The simple filter (`filter`) operation recalculates scores of objects in B within given threshold value or function based on their attributes. One can juxtaposes the filters with the selection operations in relational algebra.

The special kind of filters is normalization (`norm`) operation which is a part of the calibration procedure making scores of different q-sets compatible.

Basically, set-theoretic operations both exact and fuzzy extends by the generic fusion operations. It unions all attributes of input arguments, probably having different nature, as sets and eliminates duplicates. However, duplicates may contribute to the output score. Further, the fusion properties depends on the scoring function and the super-union operation (`superunion`) is applied (see Section 5). Whereas arguments are not required to be the same type, all distinct values of their attributes are enveloped by resulting q-set.

Finally, the aggregation operation (`aggregation`) constructs one outgoing object with surrogate identity from grouping ones by rule determined in targeting function on their attributes.

The declarative quality of operations outlined above has potential to manage both logical and physical representation of data.

5. Real-life Example

As an illustration for the building analytical system work, OpenStreetMap (OSM) data extraction and primary preprocessing can be proposed. It is pivotal for Petrozavodsk urban space investigation to manage the city's objects description in geospatial data format. More than that, mapping activity fetched from this crowd source project may provide additional statistics on community, characterizing it in implicit or explicit way.

In order to manipulate actual Petrozavodsk data from OSM servers, a replication procedure should be mount. Ordinary methods which might be found on the project Wiki-page [15] assume making a first local copy of a target area (dump) and then updating data at a certain replication interval using changes logged on the server side. It may lead previous map objects' states to be lost which is not appropriate for analytical purposes. Moreover analysis of logs of changes (changesets) is sometimes confusing.

Due to these limitations as well as comparatively small quantity of OSM objects for Petrozavodsk mapping (182211 for now), the complete copies were downloaded and then analyzed by means of the building system as outlined in experiments below.

At the first stage, appropriate querying engine settings to perform data pulling using Overpass API and primary q-sets extraction were mount. The data was replicated at $\sim$ 2-minute interval during almost 58 hours as a parallel task. It resulted in 1400 local Petrozavodsk data copies handling in auxiliary tables of the engine schema: storing, processing, removing, etc.

Basically, OSM file format (.osm) contains XML tree from instances of three types of data primitives: nodes, ways and relations with unique identities (IDs), which will be referred as OSM objects. A node primitive to specify the kind of waste system component in Petrozavodsk is shown in Listing 1.

The base set for a primary q-set consisted of an instance of OSM objects: $B = \{n \oplus w \oplus r \,|\, n \in \text{Nodes}, w \in \text{Ways}, r \in \text{Relations}\}$.

The scoring function returned as a result the value based on the index number of the dump D containing b. Assume the scores within primary query q equal to the indices of the matching dumps.

Listing 1: OSM node primitive

```
<lat>881192653</lat><lon>381984186</lon>
<tags>
  <element>waste</element>
  <element>waste_disposal</element>
</tags>
```

Let OSM object be *likely extant* if more than one out of $k = 3$ last dumps include it. Therefore, only the dumps with relative order 1, 2 and 3 are considered. All these assumption are explained by the intrinsic process of OSM modifying: while k serves more as a heuristic parameter, the weight of a later server snapshot is tend to be more significant in terms of actuality.

After the very first three dumps were loaded, the primary q-sets were composed and their scores reduced to 0.3, 0.6 and 0.9 correspondingly, within simple calibration procedure (`norm`), a super-union operation might be applied in Listing 2.

Listing 2: Extant objects detection

```
FILTER(sub.score>0.9,
 FUSION SUPERUNION(
  NORM(
   primary_qset(.osm) AS pq
   ORDER BY pq.score DESC LIMIT
        3)) sub)
```

Listing 3: Prior removing

```
DELETE FROM(
 FILTER(
  score<asOld,
  primary_qset(.osm)))
```

The scores are calculated by the Formula (1).

$$1 - \big((1 - p_{i,1}.\text{score}) \times (1 - p_{i,2}.\text{score}) \times \cdots \times (1 - p_{i,k}.\text{score})\big), \tag{1}$$

where $i \in IDs, p_{i,k}$ denotes a primary q-set, $k = 3$ was defined earlier. The base set B for the super-union q directly comprises primary q-sets.

The super-union alternative of fusion is chosen because of its nice properties. Specifically, Skimming Effect and Chorus Effect are inherent to scores evaluation of OSM objects in primary q-sets, and the super-union operation has an ability to capture these characteristics [24].

By the definition of the fusion operations (see Section 4), OSM objects in resulting q-set will differ from primary ones if and only if they are changed in at least one of evaluated copies. Meanwhile, more powerful changes detection might be provided with the class of aggregation operations.

Let function `dupls`(B, accuracy) compare objects b in primary q-sets associated with Nodes which are a base set B of some applied operation by *longitude*, *latitude* and *tag* attributes with given accuracy and greater contribution given by the values of the first two attributes. Further, it will be used in the fusion superunion Listing 4 and the aggregation specification Listing 5.

Listing 4: Duplicates detection #1

```
FILTER(dupls(attributes,
accuracy)>threshold,
 FUSION SUPERUNION(
  NORM(
   primary_qset(.osm) AS pq
   ORDER BY pq.score DESC LIMIT
        3)) sub)
```

Listing 5: Duplicates detection #2

```
FILTER(scores<>1,
 AGGREGATE(ID,
  dupls(attributes,accuracy),
  primary_qset(.osm) AS pq
   ORDER BY pq.score DESC LIMIT
        3))
```

Since the aggregation operation (see Section 4) depends on targeting aggregation functions and an attribute by which objects are grouped, different kinds of duplicates or near duplicates might be detected. For example, objects with different IDs but identical attributes are interpreted as OSM objects duplicates and should be removed. Frequently this task (conflation) is done manually, whereas it could be automated with the aid of scores obtained in the same manner as outlined above.

Table 3 shows the experiments results during which adding of 4 OSM Nodes with IDs: 2806791140; 2806793108; 2806793109 (in dump #152); 2808005201 (in dump #546) was detected by Listing 2. It shows changes in scores which were logged while sliding window of 3 dumps analysis with step size equal 1.

Also a point with geographical coordinates: $61°.80'N$, $34°.33'E$ was slightly moved during replication procedure and analysis, according to Listing 2 implementation.

Acquiring any objects with intended scores for different kinds of data manipulations is done by filters. The high-level scripts in Listings 2, 4, 5 extracts objects with scores of queries' goals, while the data flow in Listing 3 helps to remove too old primary objects from the q-set storage (threshold values, if any, are taken from the experiments).

Table 3. Fusion Superunion Scores.

ID	D151	D152	D153=D545	D546	change in scores
2806791140		$=2\rightarrow1$	$=3\rightarrow2\rightarrow1$	...	NA $\rightarrow$ 0.9 $\rightarrow$ 0.96 $\rightarrow$ 0.972
2806793108		$=2\rightarrow1$	$=3\rightarrow2\rightarrow1$	...	NA $\rightarrow$ 0.9 $\rightarrow$ 0.96 $\rightarrow$ 0.972
2806793109		$=2\rightarrow1$	$=3\rightarrow2\rightarrow1$	...	NA $\rightarrow$ 0.9 $\rightarrow$ 0.96 $\rightarrow$ 0.972
2808005201				$=2$	NA $\rightarrow$ 0.9 $\rightarrow$ ...

6. Discussion

In this paper, a preliminary stage of the ongoing research is described. While the aim of the extensive research addresses the development of a certain analytical techniques for the efficient introduction of MSW separate collection and recovery in Petrozavodsk, the present paper introduces the frameworks for urban data querying and MSWMS analysis application.

As most of the studies targeting MSW issues in Russia faced problems with data such as unavailability, inaccuracy, etc., the developed approach (or its components) along with shown solutions and applied tools might be put into practice.

An account of the preliminary study stage is the main limitation of this paper. Though only one example of data extraction from data source is considered, the approach developed allows one to extract data from other sources in a similar manner while applying various algebraic operations in queries of different depth as well as to reproduce the system building issues.

With the aid of OpenStreetMap example, results obtained indicate tools feature to tackle several functional issues during urban data preprocessing simultaneously.

Since the query language works efficiently irrespective of nature of data sources, almost the same analytical system might be mount in other cities as their uniqueness is reduced to data sources “profile”. Thus, the instruments proposed are universal for urban data management tasks.

7. Related Work

Urban data management techniques improvement while tackling real-life challenge solutions has considerable importance, which was frequently emphasized by urban researchers [2,3,18].

There is almost infinite number of different practical urban problems and tasks, where urban informatics approaches have a great potential. For instance, individuals interactions with urban centers with the aid of “travel cards in/out taps” data are studied in [20]. The platform for urban planning participation is developed in [4] and relevant Twitter messages are analyzed.

The MSWS challenge in the Petrozavodsk area is a case study for the research, the first stage of which is introduced in the present paper.

Development of query languages to facilitate complex large-scale data operation is an obvious need. Design of high-level declarative querying tools implemented on the top of existing technologies is a widespread idea among researchers and specialists [12,5,25]. Whereas the authors in [25] established the system for efficient data processing parallelism, authors in [5] focused more on the problem of domain specific entities matching.

In this paper, the instruments partly described in [12,11] are applied and improved. The main motivation for this is that analytical scenarios are translated easily to the query language, if a user is only aware of sources data model, their access details and algebra operations to perform integration and analysis.

In [18], traditional information system is designed for making different kinds of urban process more efficient, but the instruments [12] allow one to construct systems under uncertain circumstances that is inherent to Petrozavodsk.

While MSWS insufficiency is just one of the urban problems, its seriousness could not be underestimated. As explained in [9], since area of the United States and the Russian Federation is significant, their waste problems have a negative impact on the worldwide society. Comparing to Russia, the US has much more comprehensive MSW disposal practice and the complex system studies are more widespread. Investigation of disposal chain using GPRS chips incorporated in trash items is introduced [16], however this type of projects requires significant amount of resources.

Studies of MSWS in Russian cities mostly face data quality concerns [9]. Applicability of a model for cities MSWS reorganization established [19,21] is directly proportional to detailed data. While economic benefits of separate collection and recovery introduction are elucidated in a numerous works [1,22,8], serious operation obstacles and the lack of resources are also underscored [1,22,9]. The situation is more favorable in

Republic of Karelia, but researches were conducted only in Kostomuksha and not in the City of Petrozavodsk [17].

8. Conclusion

In the present paper, the first stage of study of municipal solid waste data was presented. The information system features based on the declarative query language for data extraction, integration and analysis dealing with the City of Petrozavodsk operation needs were described. The query paradigm makes design of the building information environment appropriate for urban data specifics and analysts' intentions. A major future direction of the research project is development of data source for MSW receptacles which will allow one to perform more complicated analytics and support tasks of introduction of MSW separate collection. As a part of this work, the query language improvement within the problem framework is also expected.

References

[1] Babanin I., Waste revolution. An economic assessment of MSW separate collection for St. Petersburg, Greenpeace Russia, 2006 [WWW] http://bit.ly/1oFuakZ (accessed 28.04.2014).

[2] Batty M. Big data; big issues. *Geographical Magazine of the Royal Geographical Society*, 2013, 75.

[3] Batty M. Big data, smart cities and city planning. *Dialogues in Human Geography*, 2013, 3(3), 274-279.

[4] Evans-Cowley J. S., Grifn G. Micro-participation: The role of microblogging in planning. *SSRN*, 2011.

[5] Hernández M., Koutrika G., Krishnamurthy R., Popa L., Wisnesky R. HIL: a high-level scripting language for entity integration *In Proceedings of the 16th International Conference on Extending Database Technology [EDBT], Genoa, Italy* ACM, 2013. 549-560

[6] Hemmi R. Waste management in northwest Russia. Presentation, Consulate General of Finland in St.Petersburg, 2013 [WWW] http://bit.ly/1gigv0A (accessed 1.02.2014).

[7] Ianiuk O., Novikov B. Waste Data Querying: the City of Petrozavodsk In: Haav H.-M., Kalja A., Robal T. (eds.) Databases and Information Systems: *Proceedings of the 11th International Baltic Conference [Baltic DB&IS], 8-11 June, Tallin, Estonia* TUT Press, 2014. 478, 281-288

[8] Larionov A. and IFC team. Municipal solid waste management: opportunities for Russia. Summary of key ndings, International Finance Corporation (World Bank Group), 2012 [WWW] http://bit.ly/135FQFE (accessed 15.01.2014).

[9] Liubarskaia M. Comparative analysis of contemporary Russian and American solid waste management practices, 2013 [WWW] http://bit.ly/1kYBJq8 (accessed 15.01.2014).

[10] Norden. W.A.S.T.E. Waste Awareness: Sorting, Treatment, Education, 2011 [WWW] http://bit.ly/ZZE7Bs (accessed 03.05.2013).

[11] Novikov B. From Databases to Big Data Querying, 2012 [WWW] http://bit.ly/S7RizH (accessed 10.08.2014).

[12] Novikov B., Vassilieva N., Yarygina A. Querying big data. In: Rachev B., Smrikarov A. (eds.) *Conference on Computer Systems and Technologies, CompSysTech 2012*. Ruse, Bulgaria: ACM, 2012. 1-10.

[13] Press Center of Petrozavodsk Administration. City news, 2010-2011 [WWW] http://bit.ly/1maKImc, http://bit.ly/1h7uwyP, http://bit.ly/1fegoqv (accessed 25.01.2014).

[14] Northern Dimension Environmental Partnership. Petrozavodsk solid waste management, 2009 [WWW] http://bit.ly/1cr8E1B (accessed 20.01.2014).

[15] OpenStreetMap Wiki. Main page [WWW] http://wiki.openstreetmap.org/wiki/Main_Page (accessed 28.04.2014).

[16] Phithakkitnukoon S., Wolf M. I., Offenhuber D., Lee D., Biderman A., Ratti. C. Tracking trash. *IEEE Pervasive Computing*, 2013, 12(2), 38-48.

[17] Piippo S. Proposal for municipal solid waste management for the City of Kostomuksha. Report for wp2: Urban and rural resource management, University of Oulu, 2012 [WWW] http://bit.ly/135FQFE (accessed 10.01.2014).

[18] Resch B., Mittlboeck M., Girardin F., Britter R.E., Ratti C. Live geography - embedded sensing for standardised urban environmental monitoring. *International Journal on Advances in Systems and Measurements*, 2009, 2(3), 156-167.

[19] Rodionov M., Nakata T. Design of an optimal waste utilization system: a case study in St.Petersburg, Russia. *Sustainability*, 2011, 3(9), 1486-1509.

[20] Roth C., Kang SM., Batty M., Barthlemy M. Structure of urban movements: Polycentric activity and entangled hierarchical ows. *PLoS ONE*, 2011, 6(1), e15923. doi: 10.1371/journal.pone.0015923

[21] Sliusar N., Armisheva G. Energy resources recovery on municipal solid waste disposal. *Sinks a Vital Element of Modern Waste Management, 2nd International Conference on Final Sinks, 16-18 May, Espoo, Finland.*, 2013

[22] Stupin I. From dumps to municipal solid waste tailings. *Ekspert*, 2012, 38, 88-91.

[23] Swedish-Karelian Business and Information Centre. W.A.S.T.E. project news, 2012-2013 [WWW] http://bit.ly/1kYtOZK, http://bit.ly/ObkKDN, http://bit.ly/1d19kYu (accessed 25.01.2014).

[24] Yarygina A., Novikov B., Vasilieva N. Processing Complex Similarity Queries: A Systematic Approach. In Johann E., Bieliková M., Tjoa A. M. (eds.) *Proceedings II of the 15th East-European Conference on Advances in Databases and Information Systems [ADBIS], 20-23 September, Vienna, Austria.* CEUR-WS.org: CEUR Workshop Proceedings, 2011. 789, 212-221.

[25] Zhou J., Bruno N., Wu M.-C., Larson P.-A., Chaiken R., Shakib D. SCOPE: parallel databases meet MapReduce. *The VLDB Journal*, 2012, 21(5), 611-636