FORMAL SPECIFICATION OF THE NOSQL DOCUMENT-ORIENTED DATA MODEL

Dmitriy BUY*, Sergey POLYAKOV*, Iuliia HRYSHKO**
*Department of Theory and Technologies of Programming, Faculty of Cybernetics
Taras Shevchenko National University of Kyiv, 60 Volodymyrska Street, City of Kyiv, Ukraine, 01033,
e-mail: buy@unicyb.kiev.ua, sergey.a.polyakov@gmail.com
**Department of Information Systems, Faculty of Cybernetics
Taras Shevchenko National University of Kyiv, 60 Volodymyrska Street, City of Kyiv, Ukraine, 01033,
e-mail: iuliia.hryshko@gmail.com

ABSTRACT

We built formal definition of the NoSQL document-oriented data model. Two formal data models were built. The first data model is based on sets and second one is based on multisets (bags). The special relations called subdocument and subrecord were introduced. It is proven that those relations are preorder. Also general results about the cofinal relation on the sets are given.

Keywords: *Data Model, NoSQL, Order, Preorder, Confinality.*

1. INTRODUCTION

Existing NoSQL DBMS are based on few data models. We can to talk rather about the new ideology of developing databases alternative to relational one than about a common platform constituting the ground of the NoSQL DBMS [1]. One of the most common types of NoSQL DBMS is document-oriented systems such as MongoDB [2] and CouchDB, which are based on the open standard for data representation and interchange JSON. Below we will consider formal models that describe the data structures used in document-oriented NoSQL DBMS [3], [4], [5], [6] and research some formal properties of these models.

The construction is based on a composition approach to programming [7]. At the time, it was successfully used to describe the semantics of relational databases and language SQL [8], [9], [10]. The constructed models use sets, multisets, and nominate sets as their basis [11], [12], [13].

2. DOCUMENTS INCLUSION ON SETS

Denote by 2^X_{fin} a set of all finite subsets of a set X, i.e. $2^X_{fin} = \{X' | X' \subseteq X \,\& X' - is\ finite\}$.

Let D be the set of the atomic data and V be the set of the names. Then the set of the nominate sets denoted by $D^V$is a set of finite mapping from V to D, i. e $D^V = \{A | A: V' \to D, V' \in 2^V_{fin}\}$.

Definition 1. The set of records RC (ReCord) and set of documents DC (DoCument) are constructed inductively by range. The set of records of the range 0 is coincident with D^V. Denote it by RC^0. The set of documents of range 0, denoted by DC^0, is set of all finite subsets of RC^0, i.e. $DC^0 = 2^{RC^0}_{fin}$.

Suppose records and documents of range 0, 1, …, i are defined. Then records of range *i*+1 are defined as $RC^{i+1} = \left(D \cup \bigcup_{j=0}^{i} DC^j\right)^V$. That means the value of name can be either atomic data or a document of one of the previous ranges. A document of the range *i*+1 is defined as finite set or records having range *i*+1, i.e. $DC^{i+1} = 2^{RC^{i+1}}_{fin}$.

Because of $DC^0 \subset DC^1 \subset DC^2 \subset \cdots$ is monotonically increasing sequence by constructing then it has limit $DC = \lim_{i\to\infty} DC^i = \bigcup_{i=0}^{\infty} DC^i$. By analogy, for records we have $RC = \lim_{i\to\infty} RC^i = \bigcup_{i=0}^{\infty} RC^i$. ■

Taking into account monotonic property the definition of records of the *i*+1 range can be rewritten as $RC^{i+1} = (D\ \cup DC^i)^V$.

Let's modify the definition of the range of a document and a record. Because of $DC^0 \subset DC^1 \subset \cdots$ and $RC^0 \subset RC^1 \subset \cdots$, then in the sense of previous definition the range of a document (record) is defined ambiguously: a document (record) of range *i*+1 can has lesser range.

That is why we introduce following constructions. By definition put $\widehat{DC}^0 = DC^0$, $\widehat{DC}^1 == DC^1 \backslash DC^0$, … , $\widehat{DC}^{i+1} = DC^{i+1} \backslash DC^i$, … and put $\widehat{RC}^0 = RC^0$, $\widehat{RC}^1 = RC^1 \backslash RC^0$, … , $\widehat{RC}^{i+1} == RC^{i+1} \backslash RC^i$, … for records.

It is obvious that $\{\widehat{DC}^i\}_{i=0,1,2,\ldots}$ and $\{\widehat{RC}^i\}_{i=0,1,2,\ldots}$ are splitting of the set of documents and set of records accordingly.

Definition 2. We say that index i of the set $\widehat{DC}^i$ is range of document d if $d \in \widehat{DC}^i$. The same is for records. ■

Similarly the relation of the inclusion $\subseteq$ for abstract set let's introduce relation to be subdocument for documents and relation to be subrecord for records which take into account inner structure of documents and records. Intuitively the relation to be subdocument (subrecord) means all information contained in subdocument (subrecord) also is contained in document (record). Denote by *sdoc* a relation to be subdocument and by *srec* a relation to be subrecord. The relations are introduced inductively by range.

Definition 3. Suppose $r_1, r_2 \in RC^0$. Then $r_1\ srec\ r_2$ if an only if $r_1 \subseteq r_2$. Similarly for documents of the zero range we have $d_1\ sdoc\ d_2$ if and only if $\forall r_1 \in d_1 \exists r_2 \in d_2 (r_1 \subseteq r_2)$.

Suppose those relations are defined for documents and records of the range j, $j = 0, 1, \ldots, i$. Then for records of the range i+1 the relation $r_1^{i+1}\ srec\ r_2^{i+1}$ means that the set of names of the record r_1^{i+1} is included into set of the names of record r_2^{i+1}. The values assigned to the equal names are simultaneously ether atomic data and are equal or they are documents. In second case their range is less then i+1 and document from record r_1^{i+1} must be the subdocument of the corresponding document from record r_2^{i+1}. Below $\pi_1^2 r$ is set of names of recors r (i.e. set of first components of all pairs forming binary relation r).

Formally the relation $srec$ is written in such way:

$$r_1^{i+1}\ srec\ r_2^{i+1} \Leftrightarrow \pi_1^2 r_1^{i+1} \subseteq \pi_1^2 r_2^{i+1} \wedge \forall v \Big(v \in \pi_1^2 r_1^{i+1} \Rightarrow r_1^{i+1}(v) = r_2^{i+1}(v)\ \wedge\ r_1^{i+1}(v) \in D \vee \exists j \Big(0 \leq j \leq i\ \wedge\ r_1^{i+1}(v), r_2^{i+1}(v)\ \in DC^j\ \wedge\ r_1^{i+1}(v)\ sdoc\ r_2^{i+1}(v)\Big)\Big).$$

Document d_1^{i+1} is the subdocument of d_2^{i+1} if and only if for any record $r_1 \in d_1^{i+1}$ there exists the record $r_2 \in d_2^{i+1}$ such that $r_1 srec\, r_2$. Notice that for record r_1, generally speaking, several corresponding records in document d_2^{i+1} can exist. ■

It is obvious that formal definition of $sdoc$ and $srec$ entirely corresponds to informal ideas about information including, given above.

Proposition 1. Relations $srec$ and $sdoc$ are preorder (i.e. they are reflexive, transitive, but generally speaking, not antisymmetric).■

First we shall show the reflexivity. The proof is by induction by range of records and documents.

The induction basis. For records and documents of the range zero there is nothing to prove. ▪

The step of induction. Suppose the proposition takes place for records and documents of the range i. Then prove that for records of the range $i+1$ the reflexivity take place as well, i.e. $r^{i+1}\ srec\ r^{i+1}$. Really:

- the set of names is included into itself;
- if a value of a name is atomic data then it is equal itself;
- if a value of name is subdocument then by construction its range is less then i+1. Therefore by inductive assumption it is the subdocument of itself.

For documents of range i reflexivity take place as well because of each record of range i, as shown above is the subrecord of itself.▪

Secondly we shall show the transitivity. Other words we shall prove that the implications $r_1\ srec\ r_2\ \&\ r_2\ srec\ r_3 \implies r_1\ srec\ r_3$ and $d_1\ sdoc\ d_2\ \&\ d_2\ sdoc\ d_3 \implies d_1\ sdoc\ d_3$ take place. The proof is by induction by range of documents and records.

The basis of induction. For documents and records of the range zero the transitivity takes place because of the relation to be subrecord coincides with the set-theoretical inclusion in this case. The transitivity of the relation *sdoc* for documents of the range zero is verified immediately. ▪

The step of induction. Suppose the transitivity take place for documents and records of the range j, $j = 0, 1, \ldots, i$. Then it takes place for documents and records of the range $i+1$.

Consider that for records. Suppose that $r_1^{i+1}\ srec\ r_2^{i+1}$ and $r_2^{i+1}\ srec\ r_3^{i+1}$.

From definition of $srec$ it follows that set of names of the record r_1^{i+1} is included into set of names of the record r_2^{i+1} and set of names of the record r_2^{i+1} is included into set of names of the record r_3^{i+1}. Then set of names of record r_1^{i+1} is included into set of names of the record r_3^{i+1}.

Two cases will be considered. Let the atomic data to be the value of a name v at record r_1^{i+1}. From definition it follows that the name v belongs to set of names of record r_2^{i+1} and its value is the same atomic data and that name belongs to set of names of record r_3^{i+1} and has the same value. Therefore if a name at record r_1^{i+1} has the atomic value then it has the same atomic value at record r_3^{i+1}.

Let the document d_1 to be the value of name v at the record r_1^{i+1}. Then the value of name v at record r_2^{i+1} is, generally speaking, another document d_2, and the value of name v at record r_3^{i+1} is the document d_3, and the following relations take place: $d_1\ sdoc\ d_2$ and $d_2\ sdoc\ d_3$. By construction, the ranges of the documents d_1, d_2, d_3 strictly less then i+1. Therefore by inductive assumption $d_1\ sdoc\ d_3$.▪

Now consider the documents of the range i+1.

Suppose $d_1^{i+1}\ sdoc\ d_2^{i+1}$ and $d_2^{i+1}\ sdoc\ d_3^{i+1}$. By definition $\forall r_1 \in d_1^{i+1} \exists r_2 \in d_2^{i+1} (r_1\ srec\ r_2)$ and $\forall r_2 \in d_2^{i+1} \exists r_3 \in d_3^{i+1} (r_2\ srec\ r_3)$. Then $\forall r_1 \in d_1^{i+1}\ \exists r_3 \in d_3^{i+1}$ $(r_1\ srec\ r_3)$.▪

At the same time antisymmetry doesn't take place for those relations. Really, consider following example. Suppose $d_1 = \{\{(a,1),(b,2)\},\{(a,1)\}\}$, and $d_2 = \{\{(a,1),(b,2)\}\}$. Then $d_1 sdoc\ d_2$ и $d_2 sdoc\ d_1$, but it is obvious that $d_1 \neq d_2$.▪■

Note, that relation *sdoc* is constructed by relation *srec* by logical scheme of relation of confinality [14]. Consider the common case.

Let $\langle D, \leq \rangle$ to be a set with a binary relation introduced (generally speaking, it is not needed the relation $\leq$ to be partial order).

Definition 4. The relation $\leq$ induces following relation of confinality $\trianglelefteq$ on Boolean $P(D)$ of the set D:

$$L_1 \trianglelefteq L_2 \Leftrightarrow \forall x\big(x \in L_1 \Rightarrow \exists y(y \in L_2\ \wedge\ x \leq y)\big).\ ■$$

Proposition 2. Following relations take place:

1. $\emptyset \trianglelefteq L$ for all $L \in P(D)$;
2. If relation $\leq$ is reflexive then relation of confinality $\trianglelefteq$ is reflexive too;
3. If relatione $\leq$ is transitive then relation of the confinality $\trianglelefteq$ is transitive;
4. If relation $\leq$ – is partial order then relation of the confinality orders partially the family of discrete subsets of set D (subset L is discrete, if $\langle L, \leq \rangle$ – trivial partially ordered set).■

Proofing. The clause 1 is verified directly: implication from definition of the relation of confinality is truth trivially. ▪

Clauses 2-3 are verified directly as well. ▪

Let's proof clause 4. For given $L_1 \trianglelefteq L_2$ and $L_2 \trianglelefteq L_1$, we will demonstrate that $L_1 = L_2$. Suppose x is arbitrary element such that $x \in L_1$. Then there is element $y \in L_2$ such that $x \leq y$. From $L_2 \trianglelefteq L_1$ it follows that for element y there exists element $z \in L_1$ such that $y \leq z$. So we have $x \leq y \leq z$. Therefore $x \leq z$ because of relation $\leq$ is transitive. Since $x, z \in L_1$ and $< L_1, \leq >$ is trivial partially ordered set we have $x = y$. Therefore $x \leq y$ and $y \leq x$. Since relation $\leq$ is antisymmetric we obtain $x = y \in L_2$.

Hence because of element x is arbitrary then $L_1 \subseteq L_2$. Inclusion $L_2 \subseteq L_1$ is proven in the same way.▪ ■

Conclusion 1. If initial relation $\leq$ is preorder then relation of confinality $\trianglelefteq$ is preorder as well.■

By this means the property to be preorder for relation *sdoc* is logical conclusion of the similar property *srec*.

On the subject of relation of confinality there are [15], [16].

3. DOCUMENTS INCLUSION ON MULTISETS

At the real document-oriented DBMS the records can have duplicates. The situation is similar to the tables in relation DBMS, where the rows are allowed to have duplicates. Therefore it is need to do the following refinement of the constructed data model using multisets. So let's consider the possibility to repeat records at documents.

Let's introduce some definitions of the multiset theory which are needed to construct our model, see [9], [12], [17], [18].

Definition 5. Multiset α with base U is a function $\alpha: U \to N^+$, where U is a set and $N^+ = \{1,2,\dots\}$ is the set of natural numbers without zero. ■

Here $\alpha(u), u \in U$, – number of copies (duplicates) of the base element u (multiple of the element u).

Denote by M_U all multisets with base U.

Now define the set of records RC_M and documents DC_M. The definition will be given inductively.

Definition 6. The set of records of range 0 is coincided with family of nominate sets, i.e. $RC_M^0 = D^V$.

The set of documents of range 0 is set of all finite multisets the bases of which are finite sets of records of range 0, i.e. $DC_M^0 = \bigcup_{d \in 2_{fin}^{RC_M^0}} M_d$.

Suppose the records and documents of range $j = 0, 1, \dots i$ are defined. Then records of range i+1 are defined similar to previous case, i.e. $RC_M^{i+1} = \left(D \cup \bigcup_{j=0}^{i} DC_M^j\right)^V$. The value of name can be either atomic data or document of the one of previous range. Correspondingly the documents of the range i+1 are introduced as finite multisets the bases of which are finite sets of records of range i+1, i.e. $DC_M^{i+1} = \bigcup_{d \in 2_{fin}^{RC_M^{i+1}}} M_d$.

■

Now let's redefine the relations to be subdocument (designate it as $sdoc_M$) and to be subrecord (designate it as $srec_M$). The relations are introduced inductively by range.

Definition 7. For records of the zero range the definition is the same, i.e. $r_1\, srec_M\, r_2$ if and only if $r_1 \subseteq r_2$. For documents of the zero range $d_1\, sdoc_M\, d_2$ if and only if $\forall r_1 \in U_{d_1} \exists r_2 \in U_{d_2} (r_1 \subseteq r_2)$, where U_{d_1} and U_{d_2} are the bases of documents d_1 *and* d_2 correspondingly.

Let's those relation to be defined for documents and records of the range *j*, *j* = 0, 1, … *i*. Then for records of the range *i*+1 the relation $r_1^{i+1}\, srec_M\, r_2^{i+1}$ is defined in the same way as above for exception if the values of the equal names are documents then they are in relation $sdoc_M$ but not $sdoc$.

For documents of the range *i*+1 the d_1^{i+1} is subdocument of d_2^{i+1} if and only if for any record $r_1 \in U_{d_1^{i+1}}$ there exists the record $r_2 \in U_{d_2^{i+1}}$ for which $r_1 srec_M\, r_2$. ■

Evidently for this definition the number of duplicate of records are not taken into account.

Proposition 3. The relations $srec_M$ and $sdoc_M$ are preorders.■

Proving is the same as for sets (proposition 1).■

4. EXAMPLES

Let's consider the document containing the grades of students by mathematics. To simplify the document representation we will write the pair (attribute, value) as 'attribute : value' and use square brackets to designate the multisets.

```
[{
    student_id : 0,
    class_id : 19,
    scores : [
        {
            discipline: mathematics,
            type : exam,
            score : 68.83
        },
        {
            discipline: mathematics,
            type : quiz,
            score : 39.66
        },
        {
            discipline: mathematics,
            type : homework,
            score : 81.04
        }
    ]
}
{
    student_id : 1,
    class_id : 28,
    scores : [
        {
            discipline: mathematics,
            type : exam,
            score : 23.09
        },
        {
            discipline: mathematics,
            type : quiz,
            score : 99.08
        },
        {
            discipline: mathematics,
            type : homework,
            score : 35.68
        }
    ]
}]
```

Now we will modify the document. Notably we will add a new homework grade to student with id 0 and class id 19 and add new student with id 2 and class id 27. The document will look as following:

```
[{
    student_id : 0,
    class_id : 19,
    scores : [
        {
            discipline: mathematics,
            type : exam,
            score : 68.83
        },
        {
            discipline: mathematics,
            type : quiz,
            score : 39.66
        },
        {
            discipline: mathematics,
            type : homework,
            score : 28.05
        },
        {
            discipline: mathematics,
            type : homework,
            score : 81.04
        }
    ]
}
{
    student_id : 1,
    class_id : 28,
    scores : [
        {
            discipline: mathematics,
            type : exam,
            score : 23.09
        },
        {
            discipline: mathematics,
            type : quiz,
            score : 99.08
        },
        {
            discipline: mathematics,
            type : homework,
            score : 35.68
        }
    ]
}
{
    student_id : 2,
    class_id : 27,
    scores : [
        {
            discipline: mathematics,
            type : exam,
            score : 15.23
        },
        {
            discipline: mathematics,
            type : quiz,
            score : 91.92
        },
        {
            discipline: mathematics,
            type : homework,
            score : 70.72
        },
        {
            discipline: mathematics,
            type : homework,
            score : 21.14
        },
        {
            discipline: mathematics,
            type : homework,
            score : 33.18
        }
    ]
}]
```

After such modification the old document will be the subdocument of the new one.

5. CONCLUSIONS

In this paper we constructed two data models which formalize a data structures used at document-oriented databases.

The first model is based on sets while as second one is based on multisets and allows the records to have duplicates within the scope of single document.

It is possible draw an analogy with relational database for which tables are defined either as record sets or record multisets. That impact on operation definition significantly.

Introduced relations to be subdocument and subrecord formalize our intuitive ideas about documents inclusion. They extend default relation to be subset (submultiset) on documents taking into account their inner structure. At the same time those relations are preorder that doesn't allow to speak about full analogy with relation "to be subset" ("to be submultiset").

Note also that preorder relation induces the order relation on corresponding factor-set in the usual way (see, for example [19]).

Preorder relation can be strengthened to order relation by means introducing additional restrictions. That will be considered in future works.

Now we give only one result – consequence of the proposition 2 (clause 4). Introduce the definition of the regular document: a document d is called regular if all its records have equal top level names:

$$d\text{–}regular \Leftrightarrow \forall r_1 \forall r_2 (r_1, r_2 \in d \Rightarrow \pi_1^2 r_1 = \pi_1^2 r_2).$$

Proposition 4. A subdocument relation $sdoc$ on regular documents set is order.■

The proof is based on clause 4 of the proposition 2. ■

Note the proposition is significant generalization of the following assertion of the table algebras theory: $< T, \prec >$ is partially ordered set, where T is set of all tables and relation $\prec$ is $t_1 \prec t_2 \Leftrightarrow \forall s_1 (s_1 \in t_1 \Rightarrow \exists s_2 (s_2 \in t_2 \land s_1 \subseteq s_2))$. It is based on string including relation $\subseteq$ like of confinality relation [10].

REFERENCES

[1] SADALAGE, J. P. - FOWLER, M.: NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence, Addison-Wesley Professional, p. 192, 2012.

[2] CHODOROW, K. - DIROLF, M.: MongoDB: The Definitive Guide, O'Reilly Media, p. 216, 2010.

[3] STROZZI, C.: NoSQL: a relational database management system, Available at: http://www.strozzi.it/cgi-bin/CSA/tw7/I/en (Accessed 17 March 2014).

[4] POKORNY, J.: NoSQL databases: a step to database scalability in web environment, *Proc. of the 13th International Conference on Information Integration and Web-Based Applications and Services,* pp. 278-283, 2011.

[5] STRAUCH, C.: NoSQL databases, Available at: http://www.christof-strauch/de/nosqldbs.pdf (Accessed 17 March 2014).

[6] CATTEL, R.: Scalable SQL and NoSQL data stores, *ACM SIGMOD Record*, no. 4, pp. 12-27, 2010.

[7] REDKO, V. N.: Bases of programmology, *Cybernetics and systems analysis*, no. 1, pp. 35-57, 2000 (in Russian).

[8] BUY, D. – BRONA, J. – ZAGORSKY, S. – POLIAKOV, S.: Compositional approach of the semantics of SQL, *Information Theories & Applications,* vol. 8, no. 3, pp. 133-142, 2001.

[9] BUY, D. – POLYAKOV, S.: Recursive queries in SQL and their generalization – system of recursive queries, *CSE'2010 International Scientific Conference on Computer Science and Engineering*, Košice - Stará Ľubovňa, Slovakia, pp. 252-257, Sep. 20-22, 2010.

[10] REDKO, V. N. - BRONA, J. I. - BUY, D. B. – POLYAKOV, S. A.: Relation databases: table algebras and SQL-like languages, Kiev: «Akademperiodika», p. 196, 2001 (in Ukrainian).

[11] BUY, D. – BOGATYREVA, J.: Multiset Bibliography. Methods of Multiset Lattice Construction, *Papers of 9th International Conference on Applied Mathematics*, Bratislava, pp. 407-413, Feb. 2-5, 2010, Bratislava.

[12] BUY, D. – BOGATYREVA, J.: Structure of Partially Ordered Family of Multisets, *Proceedings of CSE 2010 International Scientific Conference on Computer Science and Engineering*, Košice – Stará Ľubovňa, Slovakia, pp. 40-43, Sep. 20-22, 2010.

[13] BUY, D. B. - BRONA, J. I.: About different methods of nomimal data definition, *Bulletin of Taras Shevchenko National University. Physics-mathematics science series*, pp. 186-194, 1994 (in Ukrainian).

[14] BURBAKI, N.: Elements de Mathematique. Premiere Partie. Les Structures Fondamentals de L'Analise, Livre 1. Theorie Des Ensembles. 1958.

[15] BUY, D. B. – KACHUTA, N. D.: Properties of relation of confinality and structure of set of partially functions, *Bulletin of Taras Shevchenko National University. Physics-mathematics science series,* pp. 125-135, 2006 (in Ukrainian).

[16] BUY, D. B. – KACHUTA, N. D.: Set-theoretic constructions of whole image transform, restriction, confinality and compatibility in bases of relation database, *Intelligence System and Computer Science: International conference*, Moscow, Oct. 23-27, 2006: proceedings. Moscow, vol. 1, pp. 72-76, 2006 (in Russian).

[17] BUY, D. – BOGATYREVA, J.: Multisets: Operations, Partial Order, Computability, Applications, *Proceedings of Informatics 2011 International Scientific Conference on Informatics*, Poznava, Slovakia, Nov. 16-18, 2011.

[18] BUY, D. – BOGATYREVA, J.: Multisets: definitions, operations, basis properties, *Proceeding of 5th International conference TAAPSD'2008*, Kiev, pp. 23-26, 2008, (in Ukrainian).

[19] SKORNYAKOV, L. A.: Elements of structure theory, Second edition. – Moscow: Nauka, p.160, 1982, (in Russian).

Received November 11, 2013, accepted December 19, 2013

BIOGRAPHIES

Dmitriy Buy was born on 10.08.1958. In 1980 he graduated (MSc) with distinction at the department of Theory of Programing of the Faculty of Cybernetics at Taras Shevchenko National University of Kyiv in Ukraine. He defended his PhD (DSc) in the field of mathematical and software support of computers and networks in 1985 (2002), his thesis title was "Primitive program algebras"("Theory of compositional type program algebras and its applications" correspondently). Since 1980 he is working as a scientist and lecture at the Department of Theory and Technology of Programing, since 2011 – full professor. His scientific research is focusing on database theory, mathematical foundation of ER-model, theory of programming.

Sergey Polyakov was born on 29.11.1962. In 1985 he graduated (MSc) at the department of Theory and Technology of Programing of the Faculty of Cybernetics at Taras Shevchenko National University of Kyiv in Ukraine. He defended his PhD in the field of mathematical and software support of computers and networks in 2011; his thesis title was "Compositional semantics of SQL-like languages". Since 1990 he is working as a scientist with the Department of Theory and Technology of Programing. His scientific research is focusing on database theory, big data, machine learning. In addition, he also investigates questions related with theory of programming.

Iuliia Hryshko was born on 12.03.1985. In 2008 she graduated (MSc) with distinction at the department of

Theory and Technology of Programming of the Faculty of Cybernetics at Taras Shevchenko National University of Kyiv. She defended her PhD in the field of theory of programming in 2011; her thesis title was "Multisets Theory and its Applications". Since 2012 she is working as a lecturer at the Department of Information Systems. Her scientific research is focusing on theory of programming, databases and multisets.