УДК 621.391. 27
МРНТИ 20.53.19

SECURITY SEMANTIC DATABASE PROBLEMS

T.T. CHINIBAYEVA

International University of Information Technologies

***Abstract**: This article is a continuation of the research work [1,2]. With the development of Big Data based on semantic technologies, the problem of protecting data from unauthorized use becomes very important. The existing set of models, methods and algorithms for ensuring the security of operating systems [4,5] and relational databases [6–8] cannot be applied to semantic databases, since semantic databases (SBS) have a strong hierarchical connection between the elements and the possibility obtaining new information by users on the basis of known facts through the use of logical rules [3].*
This article discusses the well-known methods and algorithms, and based on the review, it proposes the development of algorithms to ensure the security of semantic databases.

***Keywords:** SPARQL, Semantic databases, RDF, OWL, Big Data, AC4RDF, AllegroGraph*

СЕМАНТИКАЛЫҚ ДЕРЕККӨР ҚАУІПСІЗДІГІНІҢ МӘСЕЛЕЛЕРІ

***Аңдатпа:** Бұл мақала [1,2] зерттеу жұмыстарының жалғасы болып табылады. Семантикалық технологиялар негізінде Үлкен деректерді әзірлеу кезінде рұқсатсыз пайдаланудан деректерді қорғау мәселесі өте маңызды болып табылады. Операциялық жүйелердің [4.5] және реляциялық дерекқорлардың [6-8] қауіпсіздігін қамтамасыз етудің қолданыстағы модельдері, әдістері мен алгоритмдері семантикалық дерекқорларға қолданылмайды, себебі семантикалық дерекқорлар элементтері арасында күшті иерархиялық байланыс бар және логикалық ережелерді қолдану арқылы белгілі фактілер негізінде қолданушылардың жаңа ақпаратты алу мүмкіндігі [3] ерекше.*
Бұл мақалада белгілі әдістер мен алгоритмдер талқыланып, шолуға негізделген семантикалық деректер қорының қауіпсіздігін қамтамасыз ету үшін алгоритмдердің дамуын ұсынады.

***Түйінді сөздер:** SPARQL, семантикалық дерекқор, RDF, OWL, Big Data, AC4RDF, AllegroGraph*

ПРОБЛЕМЫ БЕЗОПАСНОСТИ СЕМАНТИЧЕСКОЙ БАЗЫ ДАННЫХ

***Аннотация:** Данная статья является продолжением исследовательской работы [1,2]. С развитием Больших Данных, основанных на семантических технологиях, проблема защиты данных от несанкционированного использования становится очень важной. Существующий набор моделей, методов и алгоритмов для обеспечения безопасности операционных систем [4,5] и реляционных баз данных [6-8] нельзя применять к семантическим базам данных, поскольку семантические базы данных имеют прочную иерархическую связь между элементами и возможность получения новой информации пользователями на основе известных фактов посредством использования логических правил [3].*
В данной статье рассматриваются известные методы и алгоритмы, и на основе обзора предлагается разработка алгоритмов для обеспечения безопасности семантических баз данных.

***Ключевые слова:** SPARQL, семантическая база данных, RDF, OWL, Big Data, AC4RDF, AllegroGraph*

Overview of security approaches for semantic databases

Currently, the following methods, models and systems for controlling user access to semantic databases are already known:

• The security subsystem in the BigData repository [9], created on the basis of a model for controlling user access to named RDF-graphs.

• Model AC4RDF [10], developed based on user access control methods at the level of triplets of RDF storage.

• AllegroGraph security subsystem [11], developed on the basis of security filters.

• The RAP system (Policy-Based Access Control for an RDF Store) [16], created on the basis of the access control policy for RDF-storage.

• Methods for controlling user access to ontology [12–16].

• Control of logical rules [17].

RDF storage security model at the level of RDF graphs

In this model, user access control to RDF storage data is performed as follows:

1. All triplets are assembled into sets of triplets, which are called named graphs.

2. Each named column is assigned a security level.

3. Each user is assigned a role and permissions.

4. User U may have access and perform various operations on triplets in accordance with the security policy defined by the named graph to which these triplets belong.

This model is highly efficient when a large group of triplets is grouped in each named graph.

However, if there is only one or two statements in the named graph, the "statement level proven" model is used, which allows determining the origin of each triplet using SPARQL queries, thus you can implement a security policy for triplets.

This model is used to ensure data security in BigData RDF storage.

Model AC4RDF

The Access Control for RDF stores (AC4RDF) model implements user access control at the level of triplets of RDF storage. This model is used to ensure the security of Sesame RDF storage. This is done by checking the rights of users, as a result of which it is determined who has access rights to the RDF triplet stored in the RDF repository.

In this model, access rights are described by the owner of the RDF data using the PolicyEditor editor, which allows you to specify user access to each RDF statement or to the RDF data column stored in the RDF repository.

The overall architecture of the AC4RDF system is shown in Figure 1.

When U users send a request q to an RDF repository, the Access Control module finds information about the user account and uses the Protune Policies module to select the policy that is applied to this user request. The Rei module rewrites the request according to a specific policy. The rewritten request is sent to the RDF repository and U users can get answers to this request (Figure 2).

AllegroGraph 4.11 security subsystem

In the semantic database AllegroGraph 4.11, a security subsystem based on a security filter (filter secrutiry), which is created by the storage administrator, is used to control user access to RDF storages.

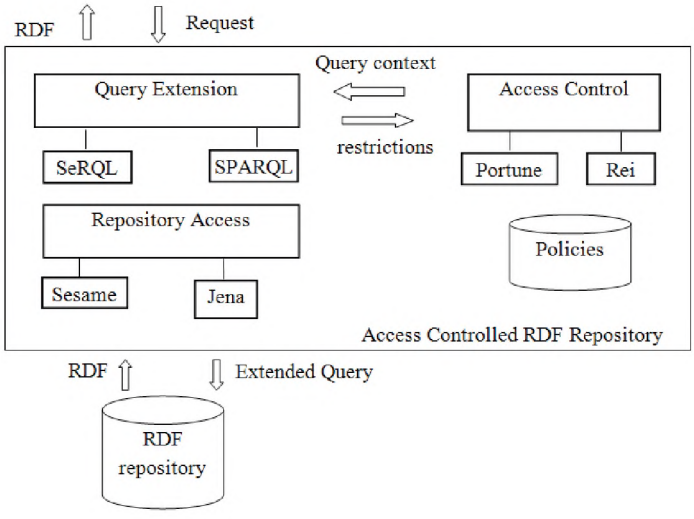

Figure 1. The overall architecture of the AC4RDF system

Figure 2. Graphical user interface for creating user security filter

The administrator has all rights to manage data and create access rights for registered users. The user is assigned a role, the value of which is selected from the set {Superuser, Start sessions, Evaluate arbitrary code, Control replication} and the rights from the set {read, write, modify or delete}.

By the security filter, the administrator assigns users access rights to any repositories, data categories (Figure 2). In addition, U users may have access only to a specific triplet or to all triplets that contain a particular predicate, subject or object.

Example of security policy: U users have the right to view all triplets containing the rec: Sarary predicate.

RAP system

In the process of working with triplets in RDF storage, user U can delete or add basic triplets that are elements of ontologies or a general scheme, therefore, the structure of the data schema (ontology) is broken. To solve this problem, a system for controlling user access to RDF storage was developed, based on policies that define user access rights.

All user actions on the repository go through the RAP system policy module to determine whether the action is "allowed" or "prohibited." In the RAP system, all triplets of metadata and access policies to them are stored in the RDF storage itself (Figure 3).

The RAP system is built on the Jena framework, in which it supports the tool for analyzing and executing simple inference on RDF, RDFS and OWL. RAP system policies are defined as rules that are used in its ontology for working with RETE. The overall architecture of this system is shown in Figure 4.

The RAP system supports the execution of various operations by users, such as adding, deleting and modifying RDF triplets in accordance with their access rights and with the correctness of the data scheme in the RDF storage.

Methods to control access to ontologies

The problem of ontology security was considered by many authors. Qin L. and Atluri V. [12] proposed a security policy scheme for controlling access to ontology concepts and their instances. Ontology concepts create security levels, and users create access levels.

Managing user access to ontologies is performed by comparing the security levels of concepts with user access levels. If the user access level is greater than the security levels of the concepts, then the users have access to the ontology concepts, therefore, they can have access to all instances of these concepts.

This system can perform control only at the level of ontology concepts, but does not understand the semantics and relations between the elements of ontologies.

Yialelis N., Lupu E. and Sloman M. [13] created a system for controlling user access to individual elements of the ontology, built on the basis of the CLP approach (constraint logic programming). This system has created a model that contains ontology and semantic data schemes. The data in this model are presented in the form of an RDF tree, on the basis of which all operations are performed that allow controlling user access to ontology elements.

In addition to the above methods and user access control systems to ontology and RDF storages, there are also other methods described in [16, 17].

Logic rule control

Currently, various methods of controlling access to logical rules have also been proposed

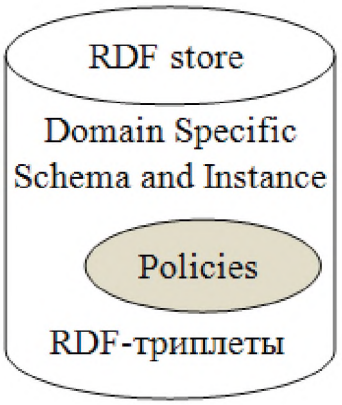

Figure 3. Data in the RDF storage

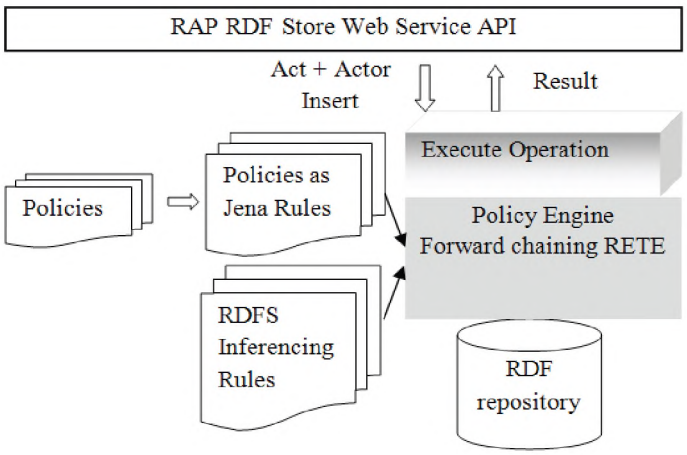

Figure 4. RAP system architecture

[17]. Basically, they are all based on the use of access levels for logical rules. In general terms, they can be described as follows: Let $DBS = \{O, M, R\}$, where O - ontologies; M - semantic metadata; $R = \{r1, ..., rn\}$ is the set of logical rules. Then the following security policy of semantic databases is used, including logical rules:

1. The set of security levels $SL = \{sl1, ..., slk\}$ is determined.

2. Each user U is given an access level $slU \in SL$ to execute logical rules.

3. Each logical rule $ri \in R$ is given an access level $slri \in SL$.

4. If $slU \geq slri$, then user U can execute the logical rules ri; otherwise, he cannot use this rule.

This method allows user U to execute logical rules in accordance with his access level, but does not guarantee that he will receive results in accordance with his access rights. This is due to the fact that in semantic databases, security levels can be specified that exceed the level of user access to the slU rules.

Proposed Algorithms for the Security of the SBD

The main features and limitations of the above subsystems, models and security methods are shown in Table 1. As a result of their analysis, we can conclude that there is no security system for semantic databases that has the following functionality:

• control of user access to individual elements of ontologies;

• control of user access to triplets and their components (subject, predicate, object);

• control of user access to RDF-graphs in the SBD;

• control of the results of logical conclusions obtained by users through the use of logical rules.

This paper proposes a security support system for working with semantic databases, which has all of the above possibilities.

This system is developed on the basis of models of control of user access to data and control of the results of logical conclusions.

The user access control model is created based on the following algorithms:

• determination of security levels of ontology and metadata elements;

• determination of security coverage (security levels of all triplets) in semantic databases;

• application of discretionary and mandatory security policies.

The model of control of the results of logical conclusions in the SBD is created based on the following methods and algorithms:

• determination of the security levels of all the findings of logical inference in the SBD;

• determination of the possibility of obtaining the results of logical conclusions between the elements;

• detection of violations of the results of logical inference in the SBD.

Table 1. Features and limitations of subsystems, models and methods for ensuring security of the SBD

Subsystems, methods, authors	Main functions	Disadvantages
Security subsystem in BigData RDF Storage	Access control at the level of named RDF graphs	No ability to control access to triplets and their components
AC4RDF	Manage user access to RDF triplet	No ability to control access to individual items
AllegroGraph security subsystem	Controlling access to a particular triplet or to all triplets that contain a particular predicate, subject or object	The system does not understand the semantics of the database. There is no possibility to control the results of logical conclusions
RAP system	Access policies are stored in RDF storage. Access control at triple level	No ability to control access to individual items. There is no possibility to control the results of logical conclusions
Subsystem L. Qin, V. Atluri	Control of access to ontology concepts and their instances	No ability to control access to attributes and ontology relationships
The subsystem N. Yialelis, E. Lupu, M. Sloman	Control user access to specific groups of ontology elements based on the RDF tree	There is no possibility to control the results of logical conclusions
Control of logical conclusions	Access control to logical rules	There is no possibility of detecting violations of the results of inference when performing logical rules

CONCLUSION

Currently, a fairly comprehensive set of tools for working with information semantics has been developed, such as: RDF - resource description language, OWL - ontology description language, SPARQL - semantic database query language, SWRL - logic rules description language.

Storage of semantic information can be implemented using semantic databases. Currently, such semantic database management systems have been developed, such as: Sesame, Oracle 11g Release, Virtuoso Universal Server.

On the basis of semantic databases, information systems are being actively created, such as, for example, semantic information portals and electronic libraries.

When working with semantic databases, two main problems need to be solved: control of user access to data and control of the results of logical deductions.

REFERENCES

1. R.Uskenbayeva, T.Chinibayeva. Algorithm for the construction of an ontology in the field of scientific knowledge//The Bulletin of Kazakh Academy of Transport and Communications named after M. Tynyshpayev ISSN 1609-1817. Vol. 107, No.4 (2018), pp. 259-266
2. R.Uskenbayeva, T.Chinibayeva. Method of extracting meta description from databases//Herald of the Kazakh-british technical university ISSN1998-6688. Vol.15, No.4 (2018), pp. 116-123
3. Hendler A. J. Handbook of Semantic Web Technologies.–Springer, 2011.– 479p.
4. Belov Ye.B. Osnivy informacionnoi bezopastnosti. – M.: Goryachiya liniya-Telecom, 2006. – 544 p.
5. Shanigin B.F. Zashita computernoi informacii. Effectivnye metody I sredstva. – Moskva: DMK Press, 2010. – 544 p.
6. Stachour P. Design of LDV: A multilevel secure relational database management system / P. Stachour, B. Thuraisingham// IEEE Transactions on Knowledge and Data Engineering 2. (1990) No3. – pp. 77–80.
7. Delugach H.S. AERIE: Database inference modeling and detection using conceptual graphs / H.S. Delugach, T. Hinke // In Proceedings of the Workshop on Conceptual Graphs. (1992) No2. – pp. 244–251.

8. ROWLBAC:representing role based access control in owl / T. Finin, A. Joshi, L. Kagal J. Niu, R. Sandhu, W. Winsborough, B. Thuraisingham // Proceedings of the 13th ACM symposium on Access control models and technologies. (2008) No2. – pp. 73–82.
9. Security model for RDF // http://www.bigdata.com/bigdata/blog/?p=307.
10. Access Control for RDF stores (AC4RDF) // http://rewerse.net/A3 /content/applications/access-control-for-rdf- storeac4rdf/index.html.
11. AllegroGraph 4.11 Security implementation //http://www.franz.com/agraph/ support/ documentation /current/ security.html#filters.
12. Qin L. Concept-level access control for the semantic web / L. Qin, V. Atluri // In ACM Workshop on XML Security. (2003) V. 11, No 1. – P. 94–103.
13. Yialelis N. Policy-based dissination of partial web-ontologies / N. Yialelis, E. Lupu, and M. Sloman // Secure Data Management 5th VLDB Workshop. (2005) No3. – P. 78–83.
14. Wang L. A logic based framework for attribute based access control/ L. Wang, D. Wijesekera, S. Jajodia // In 2nd ACM Workshop on Formal Methods in Security Engineering. (2004) No 1. – P. 110–122.
15. Reddivari P. Policy based access control for a rdf store / P. Reddivari, T. Finin, A. Joshi // In Proceedings of the Policy Management for the Web workshop, 14th International World Wide Web Conference. (2005) P. 44–47.
16. Kagal L. A policy based approach to security for the semantic web / L. Kagal, T. Finin, A. Joshi // In 2nd International Semantic Web Conference (ISWC). (2003) P. 91–96.
17. Bhavani T. Building Trustworthy semantic webs. – Francis Group, 2008. – 434p.