МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ



# МАТЕРИАЛЫ

## VI Международной молодежной научной конференции

# «МАТЕМАТИЧЕСКОЕ И ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ИНФОРМАЦИОННЫХ, ТЕХНИЧЕСКИХ И ЭКОНОМИЧЕСКИХ СИСТЕМ»

**Томск, 24–26 мая 2018 г.**

*Под общей редакцией
кандидата технических наук И.С. Шмырина*

Томск
Издательский Дом Томского государственного университета
2018

id of the discriminant analysis approach with NN technique, probabilistic NN, partial logistic NN, metaplasticity NN and hybrid NN. Possible particular methods of neural networks are feedforward neural network, multi-layer perceptron, modular neural networks, radial basis function neural networks and self-organizing network.

### 9. Model comparison

LOGR models are widely used for developing credit scoring models [3,4]. In these studies, the LOGR credit scoring models achieved better in terms of accuracy when LOGR in comparison with other models such as LR, NN, KNN and DTs. LOGR is identified as a good alternative to NN. However, it has weakness due to the model assumption, that independent variables must be linearly related to the logit of the dependent variable [4]. In some datasets, the neural networks have the highest average correct classification rate when compared with other traditional techniques, such as discriminant analysis and logistic regression, taking into account the fact that results were very close.

DTs, rule-based classifiers and any other rule extraction techniques are welcomed in the credit scoring and banking industry because of their explicit conditions in accepting/rejecting applicants so that they are easily understandable by business people in comparison with other techniques [5].

The choice of the technique is a subjective matter and depends on analyst. The most commonly used technique is NN.

### Conclusion

In this paper were considered and compared six classification techniques for credit scoring.

The result of the study demonstrates that NN classifiers perform very well. However, simple LR and LOGR also give good performances. It may indicate that most credit scoring data sets are only slightly non-linear. Although performance of other classification techniques is lower, they may compete with each other.

### REFERENCES


1. *Huang S.C., Day M.Y.* A comparative study of data mining techniques for credit scoring in banking // IEEE 14th International Conference on Information Reuse & Integration (IRI), San Francisco, CA. – 2013. – P. 684–691.
2. *Kambal E., Osman I., Taha M., Mohammed N., Mohammed S.* Credit scoring using data mining techniques with particular reference to Sudanese banks // International Conference On Computing, Electrical And Electronic Engineering (ICCEEE). – Khartoum, 2013. – P. 378–383.
3. *Louzada F., Ara A., Fernandes G.B.* Classification methods applied to credit scoring: Systematic review and overall comparison // Surveys in Operations Research and Management Science. – 2016. –V. 21. – Is. 2. – P. 117–134.
4. *Baesens B., Seow H.V., Thomas L.C.* Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research // European Journal of Operational Research. – 2015. – V. 247. – Is. 1. – P. 124–136.
5. *Sadatrasoul S.M., Gholamian M.R., Siami M., Hajimohammadi Z.* Credit scoring in banks and financial institutions via data mining techniques: A literature review // Journal of AI and Data Mining. – 2013. – V. 1. – No. 2. – P. 119–129.


## THE USE OF «IS-A»-RELATIONS IN INFORMATION TECHNOLOGY


**A.M. Babanov, E.S. Kvach**

*Tomsk State University*

babanov2000@mail.ru, kvachelena@gmail.com


### Introduction

Any sphere of human activity can not do without using «IS-A»-relations which make it possible to pass from the individual phenomena to their abstractions of different levels and vice versa, for example, «student IS-A person», «bachelor IS-A student». In order to analyze the features of «IS-A»-relations in information systems, various precedents for usage of generalized and specialized concepts in artificial intelligence, as well as semantic data models and programming the authors were considered.

## 1. «IS-A»-relations in artificial intelligence

In artificial intelligence, «IS-A»-relations are used in semantic networks. Also, they are considered to be the main elements in the network. The «IS-A» relationship establishes a type hierarchy in the network. Moreover, it is used to define the most specific semantic type [1]. A type hierarchy is a complex network of types or concepts which are organized according to levels of generality, where the concepts get more abstract as one moves up the hierarchy. In a hierarchically structured semantic net, the properties and relations of any given type can be inherited by all of its subtypes [2]. In semantic networks, «IS-A» relationships are distinguished between classes (IS-a type of) and between instances and classes (IS-an instance of) [3]. Fig. 1 shows an example of representation of «IS-A»-relations in semantic networks.

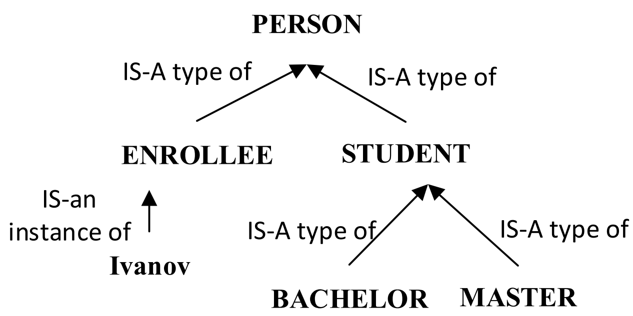


Fig. 1. Representation of «IS-A»-relations in semantic networks

Important integrity constraints are set for «IS-A»-relations: they are transitive, directed, and do not form loops. «Because of the transitivity of the relations, properties may be inherited through many levels of the type hierarchy» [4]. If every bachelor is a student and every student is a person, then every bachelor is a person, whether or not there is «IS-A»-relation explicitly saying so. «Also, we need to check that the hierarchies do not have loops and to always mark either upward or downward from a node or set of nodes, never both at once» [4].

According to Fahlman S.E. «the type hierarchy is the most important single organizational structure in the knowledge network». If information system is drawn (on paper or mentally) as a chaotic tangle of links of all kinds, it is much harder to visualize its behavior as an orderly, structured flow of events [4].

## 2. «IS-A»-relations in semantic data models

In data modeling the form of the information representation, which is comprehensive for people, it appeared together with the Enhanced Entity-Relationship Model (EER-model) in 80s of the last century [5]. In EER-model, «IS-A»-relation is referred to as «specialization». It is also used to denote the top-down process of «IS-A»- hierarchy formation with full inheritance of characteristics. The reverse process is called generalization. According to data modeling specialists, specialization is:

– «the process of identifying subclasses in the superclass, which is based on grouping unique characteristics and relationships of the subtypes» [5];
– «the process of maximizing the differences between members of an entity» [6];
– «the process of defining a set of subclasses of an entity type on the basis of some distinguishing characteristic of the entities in the superclass» [7];
– «the process of conceptual refinement to form specialized subclasses for entity sets» [8];

– «procedure in which subtypes of a more general object type are introduced to declare that specific roles are recorded only for these subtypes» [9].

Many researchers of the semantic models and the authors, who study them, use the term «specialization» to denote the process of generating subclasses. At the same time it refers to the resulting graph known as a two-level hierarchy, e.g. Terry Halpin [9], the author of Object-Role Model (OR-model), defines specialization as «procedure in which subtypes of a more general object type are introduced», and as «a subtype graph may arise in a top-down way, by specializing an object type into subtypes».

Important concepts accompanying the consideration of «IS-A»-relations are the concepts applied to the elements of these relations.

«If instances of an object type are classified into a more specific type, this specialized type is known as a subtype. A supertype is a more general form of its subtypes, and a subtype is a special form of its supertype(s).» [9].

«Superclass – an entity type that includes one or more distinct subgroupings of its occurrences, which require to be represented in a data model. Subclass – a distinct subgrouping of occurrences of an entity type, which require to be represented in a data model. The relationship between a superclass and a subclass is one-to-one (1:1) and is called a superclass/subclass relationship [6].

Consequently, the main purpose of specialization in data modeling is the possibility of grouping characteristics i.e. generalized ones rise to the level of a superclass, whereas specialized ones descend to the corresponding subclass. It should be noted that data modeling specialists [5–9] can mainly see structural and restrictive features of data classes. In addition, Terry Halpin [9], notes that specializations let make optional superclass roles mandatory for a subclass and thereby introduce additional integrity constraints. Similar to restrictions on «IS-A»-relations in artificial intelligence [4] T. Halpin argues, that specialization is transitive and any pattern of type-subtype relationships forms a directed acyclic graph: «Since no type can be a proper subtype of itself, it follows that no cycles or loops are permitted, no matter what the length of the cycle».

Additional rules for subtypes are introduced by the ER-model in Barker notation in Oracle Designer: mutually exclusive rule (each supertype instance is simultaneously an instance of one and only one subtype) and exhaustive (each supertype instance must be an instance of one of the subtypes) [10]. However, these rules only cover a full disjoint specialization. In addition to this type of specializations:

– full overlapping (each supertype instance necessarily must be an instance of one or more subtypes);
– partial disjoint (a supertype instance may be (or may not be) an instance of no more than one subtype);
– partial overlapping (a supertype instance may be (or may not be) an instance of one or more subtypes)

are also used in many data models.

Fig. 2 shows the options of representations of full disjoint specialization in semantic data models.

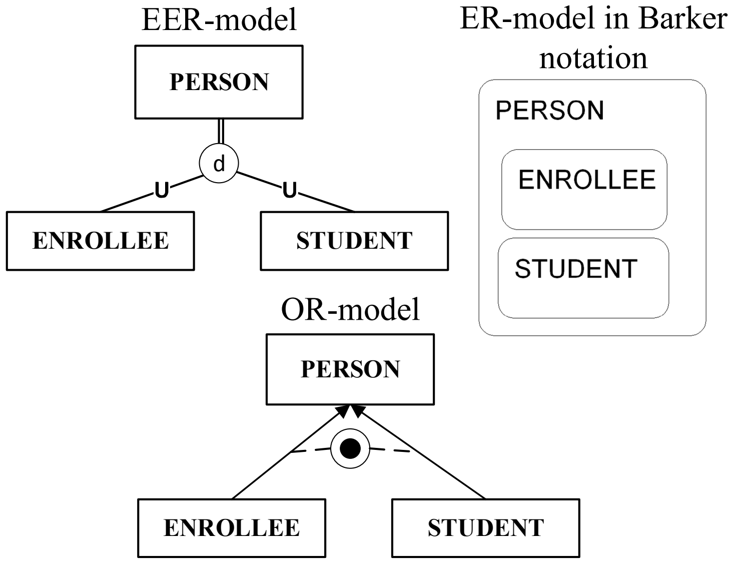


Fig. 2. Options for the representation of «IS-A»-relations in semantic data models

### 3. «IS-A»-relations in programming

In object-oriented programming (OOP) «IS-A»-relations can be found in the Unified Modeling Language (UML). The term generalization is used to denote:

- «the relationship between a class and one or more of its variations. The generalization unites classes by their general properties» [11];
- «the link between the entity of a general nature and more specific entity» [12];
- «a relationship in which a subclass inherits all the methods and fields of the superclass and can override inherited methods» [13];
- «the process associated with the identification of the common ground between the concepts and the definition of the superclass and its associated subclasses» [14].

A generalization connects generalized classes to more specialized classes, and is therefore called the inheritance relationships (class-subclass or parent-child). Generalization is called the «IS-A»-relations: for example, a student's class is-a kind of a more General entity, a class of a person [12].

In accordance with the features of the OOP, in UML, besides attributes and relations, subtypes inherit also the methods of supertype. Moreover, the implementation of methods in the subtype replaces the implementation of the same supertype method – this phenomenon is called polymorphism [12]. Experts studying UML, additionally introduce the notion of discriminator [13] or a set of generalizations [11,12]. A generalization with a name means the decomposition of a superclass into subclasses in some particular aspect called the generalization set [12]. The name of the generalization set is an enumerable attribute that indicates which aspect of the subclasses is generalized. Each generalization should be made on one aspect. For example, the generalization aspect for a *person* class is the *order for admission*. The values of the generalization set are in one – to-one correspondence with the subclasses (if there is an *order for admission*, then the *person* is a *student*, otherwise *person* is an *enrollee*). Fig. 3 shows an example of representation of «IS-A»-relations UML.

Larman K. [14] proposes two rules for subclasses: compatibility of subclass definitions (100% rule) and compatibility of multiple subclasses (IS-a rule). The 100% rule states that 100% of the superclass definition must be applied to the subclass. The subclass must 100% match its superclass in the following parameters: attributes, associations. This means that the *enrollee* and *student* subclasses must contain the *name* attribute and be associated with the *address* class. The rule «IS-A» states that all the elements of a subclass set must be elements of its superclass set. Therefore, subclass *student* should be an element of *people* set.

UML also specifies that there can be no cycles in the inheritance system: no class can be its own ancestor. The generalization relation is asymmetric: the child inherits the parent properties, but the parent does not contain any specific information about its childs [12]. Finally, the generalization is transitive and operates through an arbitrary number of hierarchy levels [11].

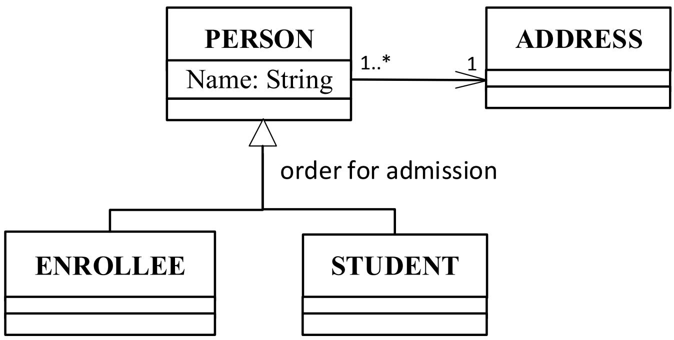


Fig. 3. Representation of «IS-A»-relations in UML

## Conclusion

«IS-A»-relations are an important element of human intellectual activity (not only scientific). This notion has been long studied by researchers, however, in each of the presented areas of knowledge different definitions of generalized and specialized concepts are given. Also different integrity constraints and design methods are proposed. Figs. 1–3 show possible variants of representation «IS-A»-relations in information technology.

After analyzing the features of «IS-A»-relations in various information technologies, the authors have concluded there is a need for stricter formalization of generalizations as well as the introduction of a new basic concept of «IS-THE»-relations in the semantic Entity-Relationship-Mapping Model [15].

## ACKNOWLEDGMENTS

The authors thank senior teacher L.V. Artamonova for help in translating the report into English.

## REFERENCES

1. U.S. National Library of Medicine, The Unified Medical Language System, UMLS Reference Manual, URL: https://www.ncbi.nlm.nih.gov/books/NBK9676/

2. *Eileen C.W.* Knowledge Representation and Metaphor. – Springer Science+Business Media Dordrecht, 1991. – 271 p.

3. *Bundy. A.* Catalogue of Artificial Intelligence Tools. – Springer-Veriag, Berlin, Heidelberg New York Tokyo, 1984. – 150 p.

4. *Fahlman S.E.* NETL: A System for Representing and Using Real-World Knowledge.– MIT Press, 1979. – 269 p.

5. *Coronel C., Morris S., Rob P.* Database Systems: Design, Implementation, and Management. – USA.: Cengage Learning, 2011. – 724 p.

6. *Connolly T., Begg C.* Database Systems: A Practical Approach to Design, Implementation, and Management, edition 4th. – Addison-Wesley, 2010. – 1400 p.

7. *Elmasi R., Navahe S.B.* Fundamentals of Database Systems, edition 6th. – Addison-Wesley, 2011. – 1201 p.

8. *Ferraggine V.E., Doorn J.H., Rivero L.C.* Handbook of Research on Innovations in Database Technologies and Applications: Current and Future Trends. – 2009. – 1124 p.

9. *Halpin T., Morgan T.* Information Modeling and Relational Databases, edition 2. – Morgan Kaufman, 2008. – 943 p.

10. *Barker R.* CASE Method: Entity Relationship Modelling. – Addison-Wesley. –1990. – 240 p.

11. *Рамбо Д., Блаха М.* UML 2.0. Объектно-ориентированное моделирование и разработка. 2-е изд. – СПб.:Питер, 2007. – 544 с.

12. *Буч Г., Рамбо Д., Якобсон А.* Язык UML. Руководство пользователя. – ДМК Пресс, 2007. – 496 с.
13. *Фаулер М., Скотт К.* UML основы. Второе издание. – Символ-Плюс, 2002. – 188 с.
14. *Ларман К.* Применение UML и шаблонов проектирования. 2-е издание Пер. с англ. – М.: Издательский дом «Вильямс», 2004. – 624 с.
15. *Бабанов А.М., Квач Е.С.* «IS-THE»-отношения в семантических моделях данных: основные понятия и разновидности // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. – 2016. – № 1(34). – С. 69–78.