



## Классификация и типизация в информатике

Статья исследует классификацию и типизацию в информационной области. Показано, что классификация и типизация строятся после создания информационной модели. Показано, что основой построения информационной модели для задач классификации является применение информационных единиц. Раскрыто содержание двух видов типизации: как обобщение знаков и как деление типов. Показано, что типизация менее строгая процедура по сравнению с классификацией. Показано, что типизация предшествует классификации.

**Ключевые слова:** информация, философия информации, классификация, типизация, информатика, информационные модели, информационные единицы



## Classification and typing in computer science

The article explores the classification and typing in the information field. The article argues that the classification and typing in computer implemented after the formation of the information model. The article argues that the basis for building information model for classification problems is the use of information units. This article describes the contents of the two types of typing: as a generalization of signs and how to divide types. The article proves that the typing less strict procedure compared with the classification. The article argues that typing precedes classification.

**Keywords:** information, information philosophy, classification, typing, computer science, information models, information units

### Введение

Классификация и типизация – две процедуры, которые служат основой систематизации и структуризации информации [1]. Традиционно считается, что типизация процедура более низкого уровня по отношению к классификации. Классификация и типизация являются объективной необходимостью, обусловленной увеличением информации накапливаемой человечеством. В процессе эволюции человеческого общества происходит исследование объектов, явлений и процессов окружающего мира. В ходе наблюдений происходит извлечение информации из информационного поля [2, 3]. В ходе исследований происходит постоянное накопление описаний объектов, явлений и процессов. Первичное описание объ-

ектов окружающего мира состоит в совокупности несистематизированных и не структурированных данных.

Чем сложнее или больше объект исследования, тем большее количество информации в его описании. Особенно остро необходимость классификации возникает при появлении проблемы больших данных [4]. Рост объемов собираемой информации и требование ее обработки и хранения делают актуальным исследование в области классификации и типизации. В работе [5] высказано предположение, что выявление закономерностей в больших массивах данных становится основным инструментом исследования и получения новых знаний. Рост объемов данных характеризует не только информатику, но и социальные системы. Все это делает актуальным исследование классификации и типизации в ин-

форматике для последующего переноса исследований в другие области.

### Объекты исследования в информатике

Объектами исследования в информатике являются среды, явления, процессы, физические и формальные объекты и модели разных классов. Это могут быть экономические, социальные, экологические, физические и другие явления. Это могут быть явления, происходящие в глобальном, национальном или региональном масштабе. Это могут быть процессы протекающие в течении нескольких лет, месяцев, недель или дней. Этими объектами могут быть проекты, технологии, системы, методы, способы, устройства и т.д. Таким образом, объекты исследования в информатике различаются по пространственному масштабу, по периоду времени исследования и анализа, по характеру существования и функционирования, по набору существенных признаков. При этом между этими разнообразными объектами существуют временные, функциональные связи, различные отношения, которые также исследуются методами информатики.

Комплексный характер объектов, изучаемых в информатике, и разнообразие областей их применения требуют их классификации и разработки различных классификаторов [6]. Для упорядочения результатов исследования в информатике необходима не только система классификаторов, но и согласованная методика классификации, которая даст возможность модернизировать существующие и создавать новые классификаторы, исключая противоречие между ними.

При использовании информационных технологий и систем возникает проблема кодирования разнообразной информации. Она также требует разработки классификаторов, обеспечивающих согласованную систему кодирования разнообразных данных. Остановимся на базовых понятиях [6].

Система классификации - совокупность правил распределения объектов заданного множества на подмножества.

Объекты классификации - предметы, понятия, свойства и другие элементы множества, для которых разрабатывается данная классификация.

Классификатор - систематизированный свод наименований классификационных группировок, объектов, признаков классификации и их кодовых обозначений.

Задачей построения классификаторов в информатике является выделение объектов классификации, создание системы правил отнесения объектов к классам. Следовательно, для проведения классификации и составления классификаторов приходится решать классификационные задачи. Классификаторы помогают систематизировать и упорядочивать большое количество информации, с которой приходится работать в

современных условиях. Актуализирует необходимость систематизации проблема «больших данных» [4] основой которой является слабая формализованность и слабая структурированность. Исходными данными для классификационного анализа могут быть результаты измерений, характеристики и признаки объектов и моделей.

В настоящее время выделяют пять подходов классификационных задач, которые определяют пять типов классификации. Для первых трех типов задач классификации воспользуемся определением, данным Кендаллом [7].

К задачам классификации первого типа относятся те, в которых исходные измерения требуется разделить на устойчивые группы. Эти задачи называют задачами классификации без учителя, кластеризации, таксономии, типизации. Этот тип классификаций основан в основном на опытных данных.

Задачи классификации второго типа характеризуются тем, что исходные данные уже сгруппированы и требуется оценить их информативность (значимость) относительно совокупности известных эталонов. Такого рода задачи встречаются при распознавании образов, при дешифрировании снимков и т.д. К задачам классификации третьего типа относят задачи разбиения. В них исходные измерения или их функции требуется разбить на устойчивые группы в зависимости от их величин.

Четвертый тип задач классификации является процедурным. В нем выполняется анализ и классификация процедур и процессов. Пятый тип задач классификации является дихотомическим [8] или структурно-аналитическим [9]. В этом типе задач производится дихотомический и логический анализ связей в системах, алгоритмах и программах. Основу такой классификации составляют методы структурного и системного анализа [10].

В информатике первая задача классификации возникает и решается при первичной обработке информации и при организации системы данных. Вторая задача классификации возникает при сборе первичных данных и при использовании информационных систем (ИС) для экспертных решений или оценок. Третья задача классификации возникает в приложениях при решении прикладных задач, например экологии, землепользования, статистики и т.п. Четвертый тип классификационных задач возникает при организации и выборе основных технологических процессов ИС. Пятый тип классификационных задач возникает при проектировании ИС как системы, при проектировании технологических процессов и при оптимизации этих процессов [11]. Строго говоря, классификационные задачи пятого типа должны предшествовать классификационным задачам четвертого типа и других типов.

Сама процедура классификации основана на том, что классифицируемый объект рассматри-

вается как совокупность наиболее важных его признаков, характеристик, по которым осуществляется анализ, сравнение и выделение классов. То есть в основу классификации в информатике используется определение информационной модели [12, 13], согласно которому информационная модель представляет собой совокупность связанных, идентифицируемых, информационно определяемых параметров, отражающих основные свойства, связи и отношения.

Все пять задач классификации могут решаться по информационной модели [14]. При этом следует иметь в виду, что информационная модель строится на объекты и процессы [15]. Таким образом, процедура классификации оказывается связанной с процедурой построения информационных моделей классифицируемых объектов.

Использование различных аспектов рассмотрения моделей позволяет выбирать различные критерии построения и сравнения моделей. Современные информационные модели строятся на основе системного подхода. Основу современных информационных моделей составляют информационные единицы [16, 17], которые являются аналогами элементов сложной системы.

Содержание элементарной информационной модели включает следующие группы характеристик, составляющие ее описание: правила построения и применения модели, имя моделируемого объекта, свойства объекта, значения свойств.

Одно из противоречий описания моделей [13] заключается в стремлении с одной; стороны, отразить общие свойства класса объектов, с другой - индивидуальные признаки более узкого подкласса и, наконец, индивидуальные признаки конкретного объекта. Таким образом, уже на стадии построения модели существует связь между ее построением и классификацией. Поэтому характеристика "имя моделируемого объекта" более точно заменится характеристиками "принадлежность к классу" и "индивидуальные свойства объекта" на основе дальнейшего анализа.

Противоречивость между общими и индивидуальными свойствами моделей приводит к разделению моделей на два класса: сильно типизированные и слабо типизированные [18]. Именно этот аспект рассмотрения и используется в дальнейшем для построения классификаторов.

Сильно типизированные - это модели, в которых предполагается, что большинство данных удовлетворяют неким условиям и ограничениям и могут быть отнесены к узкому подклассу (типу). Если исходные данные нельзя отнести к одному типу, то их можно с помощью искусственных приемов (введением дополнительных условий или ограничений) отнести к набору типов.

Примером сильно типизированных данных служат координатные данные. Слабо типизированные модели это модели, в которых данные разнородны по формату, структуре. Они, в общем, слабо связаны условиями относительно

известных типов. Примером слабо типизированных моделей могут быть описательные характеристики, текстовые файлы.

Сильно типизированные модели эффективны при компьютерной обработке потоков данных. Слабо типизированные модели обеспечивают интеграцию категорий данных. Предельная возможность использования таких моделей обеспечивается исчислением предикатов.

Проанализируем группы характеристик элементарной модели: правила построения и применения модели, имя моделируемого объекта, свойства объекта, значения свойств. В информатике основой моделирования является информационное моделирование, которое применяет информационные единицы. Определим информационные единицы, отражающие группы характеристик элементарной модели. Для "значения свойств" информационной модели используют информационную характеристику "знак"[18, 19].

Знак - элементарная информационная единица, являющаяся реализацией свойств объекта в заранее заданной, структурно организованной знаковой системе. Например:

1). В знаковой системе целых чисел знаками будут целые числа типа 1,2, 10, 101 и т.д. Но в этой системе знаком не будет являться дробные числа.

2). В знаковой системе вещественных чисел в качестве знаков будут выступать числа типа 1, 1.3, 5.356 и т.д.

3). В системе русского алфавита знаками будут буквы алфавита а, б, т и т.д.

4). В системе латинского алфавита знаками будут латинские буквы X, Y, Z и т.д., но не русские П, Б и т.п.

Знак можно рассматривать как элемент модели, одновременно являющийся элементом какой-то существующей классификации. Присвоения знака модели дает возможность осуществлять ее частичную классификацию по значениям, т.е. "по знакам". Таким образом, анализ "значения свойств" модели позволяет осуществлять ее частичную классификацию, т.е. классификацию объекта, который данная модель отображает. Эту классификацию назовем "классификацией по знакам".

Для группы характеристик модели "качественные свойства" используют информационную характеристику "тип". Тип - Совокупность моделей или объектов, объединенная общим набором признаков или класс подобных знаков [19]. Тип, как и знак, может принадлежать некой известной классификации. Однако возможны случаи, когда набор признаков данного типа не дает возможность отнести его к известной классификации. Это имеет место, если модель слабо типизирована.

Тип более мягкая характеристика в сравнении с классом. Класс результат классификации,

которая имеет классификационное основание и является согласованной системой. Тип является более произвольной системой. Типизацию может осуществлять любой исследователь по своим критериям без согласования такой типизации с другими. Такая типизация является условной и не согласованной. Тем не менее, такая типизация может приводить к открытиям новых закономерностей.

В приведенном выше примере для первой рассмотренной знаковой системы типов будет тип "целый". Для второй группы знаков типом системы будет тип "вещественный". Для третьей и четвертой групп типов данных будет "символьный" или "текстовый".

Типизация - объединение данных по набору заданных признаков или разбиение множества на подмножества по меньшему числу типов.

Знак можно рассматривать как реализацию типа. Тип можно рассматривать как обобщение совокупности знаков или типов. Следовательно, знак представляет индивидуальные свойства модели, а тип ее общие свойства. Соответственно процедуры реализации знака и типизация являются противоположными. В информатике типизацию осуществляют двумя путями: обобщение совокупности знаков, деление общего типа на более мелкие. Типизация предшествует классификации и может служить основой для нее.

Эти процессы показаны на рис.1 и 2. Типизация как обобщение знаков используется для построения "типизация по знакам" рис.1. Если модель слабо типизирована, выполняется разбиение на типы рис.2.

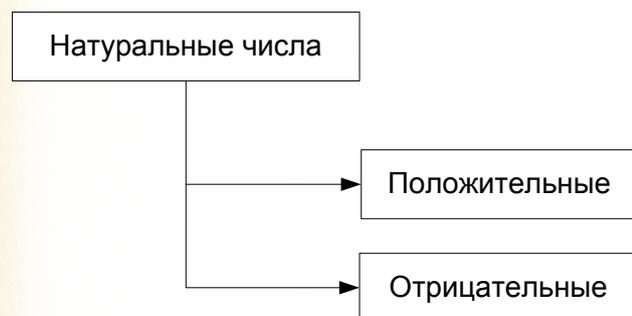


Рис. 1. Типизация по обобщению знаков



Рис. 2. Разбиение типов

Это дает возможность в дальнейшем выделить такой тип характеристик модели "свойства объекта", который затем можно отнести к известному классу. Эта классификация называется "классификацией по типам".

Таким образом, классификация в информатике включает построение модели объекта, преобразование ее в информационную модель, анализ свойств объекта и разбиение их на типы, анализ значения свойств объекта и их типизацию. Имя объекта уточняется после классификации двух отмеченных групп, что дает возможность осуществить идентификацию модели объекта в соответствии с проведенной классификацией.

После этого уточняются правила построения модели объекта для того, что бы ее можно было эффективней относить к известным классам и проводить ее анализ. Все это в итоге повышает эффективность обработки, анализа и хранения информации в информатике.

Следует отметить, что в зависимости от метода выбора признаков может меняться типизация как разбиение на типы или обобщение знаков в виде типа. Эти процессы показаны на рис.2. на нем те же исходные множества, что и на рисунке 1 типизированы по другим признакам.



Рис. 3. Альтернативная типизация

Это дает возможность выбирать разные аспекты типизации, анализа данных и соответственно обеспечивает выбор разных классификаций.

## Заключение

Классификация в информатике создается на основе построения и анализа информационных моделей. В свою очередь, информационные модели должны формироваться на основе применения информационных единиц. Классификация - строгая процедура [20], которая обладает полнотой целостностью и переносимостью. Классификации предшествует типизация, которая является менее строгой и может быть условной. Переход от типизации к классификации осуществляется на основе системного и дихотомического и логического анализа.

1. Бондур В.Г., Кондратьев К.Я., Крапивин В.Ф., Савиных В.П. Мониторинг и предсказание природных катастроф // Проблемы окружающей среды и природных ресурсов. 2004. №9. С.3-8.
2. Цветков В.Я. Естественное и искусственное информационное поле // Международный журнал прикладных и фундаментальных исследований. 2014. №5. Ч.2. С.178 -180.
3. Бондур В.Г. Информационные поля в космических исследованиях // Образовательные ресурсы и технологии. 2015. №2 (10). С.107-113.
4. Майер-Шенбергер В., Кукьер К. Большие данные: Революция, которая изменит то, как мы живем, работаем и мыслим. – Манн, Иванов и Фербер, 2014. 240 с.
5. The Fourth Paradigm: Data-Intensive Scientific Discovery, 2009, URL: <http://research.microsoft.com/enus/collaboration/fourthparadigm>
6. Саков А.А. Унификация управленческой документации и общесоюзные классификаторы. М.: Экономика, 1982.
7. Кендалл М. Дж., Стюарт А. Многомерный статистический анализ и временные ряды. М.: Наука, 1976. С. 437-473.
8. Tsvetkov V.Ya. Dichotomous Systemic Analysis. Life Science Journal. 2014. 11(6). pp586-590.
9. Елсуков П.Ю. Управление с использованием информационных методов // Государственный советник. 2015. №2. С. 29-33.
10. Берталанфи фон Л. Общая теория систем – критический обзор. / В кн. Исследования по общей теории систем. М.: Прогресс, 1969. С. 23 -82.
11. Соловьёв И.В., Майоров А.А. Проектирование информационных систем. Фундаментальный курс / Под ред. В.П. Савиных. М.: Академический проект, 2009.
12. Поляков А.А., Цветков В.Я. Прикладная информатика. Учебно-методическое пособие для студентов, обучающихся по специальности «прикладная информатика» (по областям) и другим междисциплинарным специальностям: В 2-х частях: / Под общ.ред. А.Н. Тихонова. М.: МАКС Пресс. Том 1. 2008. 788 с.
13. Цветков В.Я. Социальные аспекты информатизации образования // Международный журнал экспериментального образования. 2013. № 4. С. 108-111.
14. Бондур В.Г., Савин А.И. Принципы моделирования полей сигналов на входе аппаратуры ДЗ аэрокосмических систем мониторинга окружающей среды // Исследование Земли из космоса. 1995. № 4. С.24-34.
15. Цветков В.Я. Информационные модели объектов, процессов и ситуаций // Дистанционное и виртуальное обучение. 2014. №5. С.4-11.
16. Ozhereleva T. A. Systematics for information units // European Researcher, 2014, Vol.(86), № 11/1, pp. 1894-1900. DOI: 10.13187/er.2014.86. 1900
17. Tsvetkov V. Ya. Information Units as the Elements of Complex Models // Nanotechnology Research and Practice, 2014, Vol.(1), № 1, p.57-64.
18. Цветков В.Я. Геоинформационные системы и технологии. М.: Финансы и статистика, 1998. 288 с.
19. ЦикритзисД., Лоховски Ф. Модели данных. М.: Финансы и статистика, 1986. 344 с.
20. Аржененко Н.И., Бондур В.Г., Классификация облачных форм по пространственным спектрам изображений // Оптика атмосферы и океана. 1988. №11. С. 38-45.

#### **Информация об авторе**

**Елсуков Павел Юрьевич**

(Россия, Иркутск)

Кандидат технических наук, старший научный сотрудник. Институт систем энергетики им. Л.А. Мелентьева Сибирского отделения Российской академии наук (ИСЭМ СО РАН)

#### **Information about the author**

**Elsukov Pavel Yurevich**

(Russia, Irkutsk)

PhD in Technical Sciences. Senior Research Associate Federal State Institution of Science Institute of Energy Systems Institute. LA Melentyeva Siberian Branch of the Russian Academy of Sciences (ESI SB RAS)