

Copyright © 2014 by Academic Publishing House *Researcher*

Published in the Russian Federation
 European Researcher
 Has been issued since 2010.
 ISSN 2219-8229
 E-ISSN 2224-0136
 Vol. 86, No. 11-1, pp. 1894-1900, 2014

DOI: 10.13187/er.2014.86.1894
www.erjournal.ru



UDC 004.8 004.9

A Systematics for Information Units

Tatiana A. Ozhereleva

Moscow State University of Geodesy and Cartography (MIIGAiK), Russian Federation
 Senior Lecturer
 E-mail: ozerotana@yandex.ru

Abstract

This article examines the use of information units in the information sciences. The author illustrates the multi-aspect nature of applying information units; examines examples of information units; illustrates the criteria for subsuming information units under different groups; divides information units into different groups based on application and reflection of objective reality; describes issues in the theory of information units.

Keywords: information; information model; linguistics; computational linguistics; information technology; information units; information construction.

Введение

Проблема информационных языков существует в науке достаточно давно. Никого не смущает термин «язык карты» [1] или «язык радиотехнических схем». Однако до сих пор нет определенного ответа на вопрос, что называют информационными языками [2] и какую роль играют единицы этих языков. Упрощенно считают, что всякий язык имеет алфавит, слова, синтаксис и прагматику. С этим можно условно согласиться. Слово можно рассматривать как носитель языка и средство передачи смысла. Но существуют разные смыслы, в том числе и такие которые в слово не входят. Есть основание говорить о носителе смысла языка. Этот носитель смысла уместно назвать семантической информационной единицей [3]. Каждый язык имеет структуру и структурные элементы. Уместно назвать структурный элемент языка структурной информационной единицей [4]. Таким образом, уже в первом приближении возникает два типа информационных единиц. Это требует анализа и систематизации информационных единиц.

Области применения информационных единиц. Прежде чем приступить к систематизации, дадим краткий обзор применения информационных единиц. Первой задачей исследования является систематизация многообразия групп информационных единиц. Каждая из групп информационных единиц, решает свои проблемы и задачи, имеет свои средства, получает особые продукты. Однако, все они, в силу независимого становления, не обобщены как особые информационные конструкции, не организованы в одну систему, и это создает смешение проблем и результатов анализа. К этому надо добавить, что собственно теоретические исследования составляют сейчас в сфере наук об информации незначительную часть в силу господства идеологии обработки информации.

Исследование многообразия требует анализа эмпирического материала. Но при этом важна другая, не теоретическая, а методологическая сторона проблемы: необходимо показать принципы и общую схему подобного анализа.

С позиций лингвистики, в языке информационная единица анализа — лингвистическая единица речи или элемент содержания, служащие в тексте индикатором интересующих исследователя явления [5]. Дополнительно к лингвистике существуют паралингвистические информационные единицы [6]. В картографии применяют картографические информационные единицы — условные знаки [1]. Информационные единицы, как и многие другие термины в области наук об информации, могут быть полисемическими. Это находит отражение в представлении их как многозначных информационных единиц [7]. С позиций структуры рассматривают информационные единицы [8] и микро информационные единицы [9]. Причем в работе [9] информационная единица рассматривается с двух позиций как элемент структуры и как элемент информационного поиска. Коммуникационные единицы [10, 11] также в сущности являются информационными единицами. Информационные единицы рассматривают как статистические информационные единицы [12] при проведении кластеризации или автоматической классификации в системах анализа текстов. В аспекте логического анализа информационные единицы рассматривают как составляющие фактов [13], которые человек получает в порядке познания окружающей действительности. В информационных поисковых системах информационные единицы применяют как логические единицы поиска информации [14]. Как логические единицы информационные единицы применяют также при анализе информационных систем [15]. В работе [16] рассматривается информационная единица как единица доступа к базе данных.

Если исходить из эмпирической схемы образования информационных единиц, то нужно констатировать, что никакое положение, полученное из анализа одних информационных единиц или сколь угодно большого их числа, не дает пока теорию информационных единиц. Необходимо строго научным образом исследовать все компоненты деятельности, участвующие в создании информационных единиц, и все варианты деятельности, создающие, если это возможно, разные виды их.

Группы информационных единиц. В настоящее время информационные единицы представляют собой совокупность групп единиц, применяемых в различных направлениях. Информационные единицы делят на простые и составные [17].

Рассмотрим наиболее часто используемые группы лингвистических информационных единиц, которые в области наук об информации называют семантическими информационными единицами [3]. Особенность современных исследований в области информации состоит в углубленном изучении смысловой стороны информации. Главным критерием наличия содержательности информации (и в информационных единицах соответственно) Л. Флориди считает истинность информации [18].

Семантические информационные единицы — это единицы, рассматриваемые в аспекте семантической содержательности. Выделяют следующие семантические информационные единицы: слово, предложение, фраза. Это выделение обусловлено разными смыслами этих единиц. Главный критерий выделения информационных единиц — неделимость по заданному критерию.

Слово — простая информационная единица, обладающая неделимостью по сигнификативному смысловому признаку. В тексте слово — предельная смысловая составляющая, способная непосредственно соотноситься с предметом отражения и указывать на него; вследствие этого слово приобретает определенные смысловые свойства. Слово минимальная единица информации, имеющая смысловое значение.

Предложение — составная информационная единица, обладающая неделимостью по предикативной смысловой совокупности связанных слов и выражающая законченную мысль. Предложение — представляет собой предикативное словосочетание, поэтому имеет предикативное смысловое значение. Предложение — предикативно неделимая информационная единица.

Фраза (фразеологическая единица) — составная информационная единица, обладающая ассоциативной неделимостью по связанности предложений и выражающая законченную мысль. *Фраза* — информационная единица сообщения, обладающая

максимальной смысловой содержательностью. Для фразы и предложения смысловое содержание дополняется контекстом.

В таблице 1 приведены основные семантические информационные единицы (СИЕ) с указанием критерия их делимости и основного значения. Термин основное значение обусловлен внутренней неоднозначностью информационных объектов и полисемией понятий, вытекающей из этого [11].

Таблица 1

Делимость информационных единиц

Часть сообщения	Критерий делимости	Основное значение
Слово	Смысл свободная форма	Сигнификативное
Предложение	Смысл связанное словосочетание	Предикативное
Фраза	Фразеологическая связь	Ассоциативное

От лингвистических информационных единиц можно перейти к графическим информационным единицам. На рис.1 приведены графические информационные единицы, образованные фигурами. В теории автоматизированного проектирования их называют «примитивами», поскольку они являются простейшими и составляют основу для создания более сложных фигур.

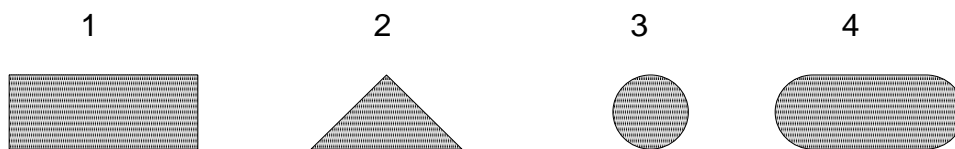


Рис. 1. Графические информационные единицы – фигуры

Все фигуры достаточно просты и нет смысла их пояснять. Графические информационные единицы, приведенные на рис 1, представляют собой элементарные геометрические образы, которые изучают в аналитической геометрии.

На рис. 2 приведена еще одна совокупность примитивов, но уже образованная линиями

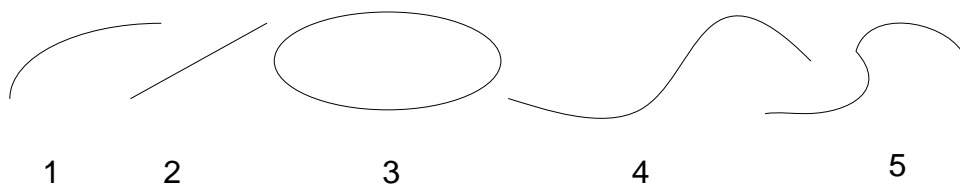


Рис. 2. Графические информационные единицы – линии

Графические единицы на рис. 1 имеют характеристику – площадь, которую не имеют графические единицы – линии на рис. 2. Можно дать характеристику примитивам на рис. 2: 1 – унимодальная кривая; 2 – отрезок прямой; 3 – замкнутый эллиптический контур (подчеркиваем контур, а не фигура) Фигуры выделяем штриховкой; 4 – произвольная гладкая кривая; 5 – Кривая со скачком производной (не гладкая кривая). Из графических информационных единиц на рисунке 2 строят любые кривые на чертежах, планах и картах.

В топологии также существуют свои информационные единицы. Они приведены на рис. 3.

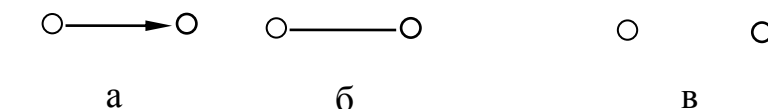


Рис. 3. Примеры топологических информационных единиц

Рис 3а описывает информационную топологическую единицу – ориентированный граф.
 Рис 3б описывает информационную топологическую единицу – не ориентированный граф.
 Рис. 3 в описывает информационную топологическую единицу – вершину или узел (их 2).

Можно для сравнения взять еще одну область – геоинформатику. На рис. 4 приведены основные графические информационные единицы геоинформатики, образованные традициями деления в этой науке.

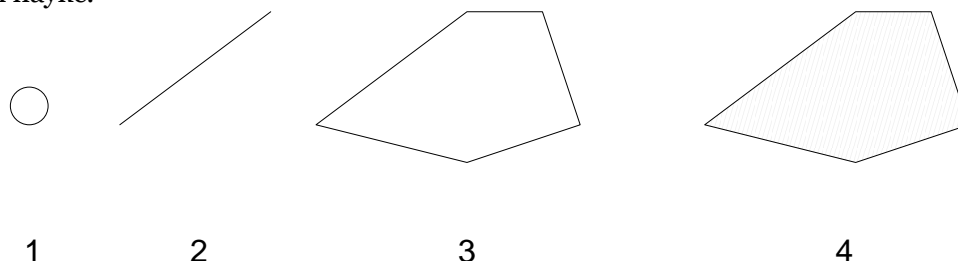


Рис. 4. Информационные единицы в геоинформатике

Интерпретация этих единиц следующая: 1 – точечный объект; 2 – линейный объект, не замкнутая линия; 3 – линейный объект, замкнутая линия, который в геоинформатике называют контур; 4 – площадной или ареальный объект.

Обращает на себя внимание сходство по форме объектов 3 и 4. Формы у них одинаковые (это сделано специально), но качества разные. Для 4 существует дополнительная характеристика площадь. Для 3 только периметр.

Обсуждение

Информационные единицы – это единицы, которые содержат порции информации или характеризуют содержание порции информации. Как базовые элементы теории, информационные единицы (ИЕ) обладают свойством неделимости по какому-либо признаку [3]. Информационные единицы служат основой построения сложных: языковых описаний, информационных конструкций [19] или информационных коллекций [20]. Как многие информационные понятия, информационные единицы (ИЕ) являются полисемическим, многоаспектным понятием. Поэтому для разграничения и уточнения видов информационных единиц необходимо указывать аспект их рассмотрения и область применения.

В аспекте структуры ИЕ выделяют составные и простые информационные единицы. Простые ИЕ не включают в свой состав другие единицы. Составные информационные единицы включают в свой состав другие информационные единицы. Например, информационная единица «предложение» включает информационные единицы «слова» [9]. По этой причине для многих составных информационных единиц имеет место характеристика – структурная вложенность. Структурная вложенность информационных единиц – это не структура, а отношение иерархии компонент единицы и ее окружения.

Для многих составных информационных единиц имеет место характеристика – окружение информационной единицы. Окружение информационной единицы – это другие, связанные с ней информационные единицы и характеристики, необходимые для однозначной интерпретации информационной единицы и ее информационной определенности. Информационное окружение единицы проявляется при ее непосредственном использовании. Например, информационным окружением информационной единицы «слово» в предложении или во фразе, будут все связанные с этим словом символы и другие слова, а также такие информационные характеристики как позиция слова и вид его написания.

В аспекте применения выделяют: формальные, логические, физические – информационные единицы. Примером физической информационной единицы являются бит, байт, кластер. Бит (байт) характеризует информационную емкость носителя информации, а не количество информации. Эти единицы являются аналогами физических единиц измерения емкости литр, кубический сантиметр и др. Кластер характеризует физическую единицу дисковой памяти. Эти характеристики применяют в компьютерных технологиях.

Примером формальной информационной единицы являются: клетка таблицы, реквизит,

элемент базы данных.

В аспекте отражения свойств внешнего мира выделяют: субстанциональные, процессуальные, атрибутивные и комбинированные – информационные единицы.

Субстанциональные информационные единицы характеризуют сущности, процессуальные – процессы, а атрибутивные единицы описывают свойства. Субстанциональные информационные единицы можно рассматривать как элементарные описания объектов, фактов, явлений – окружающего мира. Графические информационные единицы являются субстанциональными, если они описывают сущности.

Примером атрибутивной информационной единицы является реквизит в базах данных. Реквизиты – логически неделимые элементы, соотносимые с определением свойств отображаемого объекта или процесса. Графические информационные единицы являются атрибутивными, если они описывают свойства или отношения. Например отношение принадлежности ISO или АКО.

Процессуальная информационная единица – это единица, которая описывает динамику: процесс или его часть. Примером процессуальных информационных единиц являются транзакция и бод, битрейт. Транзакцией называют неделимую последовательность операций манипулирования данными в БД. Это свойство информационной единицы. Графические информационные единицы являются процессуальными, если они описывают процессы или информационные потоки. Например, стрелки в алгоритмах показывают процесс обработки информации.

Заключение

Информационные единицы в технологическом аспекте могут быть рассмотрены как особые информационные модели. Как элементы языка информационные единицы служат его основой только в картографии [1]. Количество условных знаков, которые образуют информационный язык в этой области или язык карт составляет более 10 000 единиц. Однако и широта описания этого языка масштабная и всеобъемлющая. В других областях информационные единицы представлены группами и часто в развитии тех информационных технологий, которые они обслуживают. Кроме того, потребность в информационном языке для отдельных технологий не существует. В них представлены группы единиц, предназначенные для решения локальных практических задач, а не для решения теоретических задач. Потребуется разработка какой-то теоретической базы для объединения информационных единиц в виде теоретической идеи. Пока общим для информационных единиц является критерии неделимости, порции действия и или порции содержательной информации. Еще одним объединяющим принципом для информационных единиц является принцип структурирования информационных коллекций. Поэтому эту область следует считать технологической и ожидать радикальных теоретических разработок. На наш взгляд перспективным направлением развития теории информационных единиц является развитие теории семантического окружения. Оно дает через среду анализировать информационные единицы и лишить их обособленности.

Примечания:

1. Лютый А.А. Язык карты: сущность, система, функции. М.: ГЕОС, 2002, Изд. 2-е. 2002. 327 с.
2. Москович В.А. Информационные языки. М.: Наука, 1971.
3. Tsvetkov V. Ya. Semantic Information Units as L. Floridi's Ideas Development // European Researcher, 2012, Vol.(25), № 7, p. 1036-1041.
4. Цветков В.Я. Информационные единицы как средство построения картины мира // Международный журнал прикладных и фундаментальных исследований. 2014. (Часть 4). № 8. С. 36-40. URL: www.rae.ru/upfs/?section=content&op=show_article&article_id=5690 (дата обращения: 01.08.2014).
5. Денисенко В.Н., Чеботарева Е.Ю. Современные психолингвистические методы анализа речевой коммуникации. М.: РУДН, 2008. 258 с.
6. Цветков В.Я. Паралингвистические информационные единицы в образовании // Перспективы науки и образования-2013. №4. С. 30-38.
7. Hill, Frederick C., and Alan Treibitz. "Method for presenting information units on multiple presentation units." U.S. Patent No. 6,091,408. 18 Jul. 2000.

8. Tsvetkov V.Ya. Information objects and information Units // European Journal of Natural History. 2009. № 2. p 99.
9. Li, Xiaoli, et al. "Using micro information units for internet search." Proceedings of the eleventh international conference on Information and knowledge management. ACM, 2002.
10. Bienvenu, Jacques, et al. "Apparatus and method for transferring information units between processes in a multiprocessing system." U.S. Patent No. 4, 394, 725. 19 Jul. 1983.
11. Цветков В.Я. Информационные единицы сообщений // Фундаментальные исследования. 2007, №12. с. 123-124.
12. Huber, D. (1989, May). A statistical approach to the segmentation and broad classification of continuous speech into phrase-sized information units. In Acoustics, Speech, and Signal Processing, 1989. ICASSP-89. 1989 International Conference on (pp. 600-603). IEEE.
13. Kratzer A. Facts: Particulars or information units? //Linguistics and philosophy. 2002. T. 25. № 5. С. 655-670.
14. Tajima K. et al. Discovery and Retrieval of Logical Information Units in Web // WOWS. 1999. С. 13-23.
15. Tsvetkov V.Ya. Logic units of information systems // European Journal of Natural History. 2009. № 2. p. 99-100.
16. Nandi A., Jagadish H.V. Qunits: queried units in database search // arXiv preprint arXiv: 0909.1765. 2009.
17. Tsvetkov V. Ya. Information Units as the Elements of Complex Models // Nanotechnology Research and Practice, 2014, Vol.(1), № 1. P. 57-64.
18. Luciano Floridi In defence of the veridical nature of semantic information // European Journal of Analytic Philosophy. 2007. Vol. 3. No. 1. p. 31- 41.
19. Tsvetkov V. Ya. Information Constructions // European Journal of Technology and Design, 2014, Vol.(5), № 3. P. 147-152.
20. Barker K. J. et al. Information collection architecture and method for a data communications network : пат. 5375070 США. 1994.

References:

1. Lyutyi A.A. Yazyk karty: sushchnost', sistema, funktsii. M.: GEOS, 2002, Izd. 2-e. 2002. 327 с.
2. Moskovich V.A. Informatsionnye yazyki. M.: Nauka, 1971.
3. Tsvetkov V. Ya. Semantic Information Units as L. Floridi's Ideas Development // European Researcher, 2012, Vol.(25), № 7, p. 1036-1041.
4. Tsvetkov V.Ya. Informatsionnye edinitsey kak sredstvo postroeniya kartiny mira // Mezhdunarodnyi zhurnal prikladnykh i fundamental'nykh issledovaniy. 2014. (Chast' 4). № 8. С. 36-40. URL: www.rae.ru/upfs/?section=content&op=show_article&article_id=5690 (data obrashcheniya: 01.08.2014).
5. Denisenko V.N., Chebotareva E.Yu. Sovremennye psikholingvisticheskie metody analiza rechevoi kommunikatsii. M.: RUDN, 2008. 258 s.
6. Tsvetkov V.Ya. Paralingvisticheskie informatsionnye edinitsey v obrazovanii// Perspektivy nauki i obrazovaniya-2013. №4. S. 30-38.
7. Hill, Frederick C., and Alan Treibitz. "Method for presenting information units on multiple presentation units." U.S. Patent No. 6,091,408. 18 Jul. 2000.
8. Tsvetkov V.Ya. Information objects and information Units // European Journal of Natural History. 2009. № 2. p 99.
9. Li, Xiaoli, et al. "Using micro information units for internet search." Proceedings of the eleventh international conference on Information and knowledge management. ACM, 2002.
10. Bienvenu, Jacques, et al. "Apparatus and method for transferring information units between processes in a multiprocessing system." U.S. Patent No. 4, 394, 725. 19 Jul. 1983.
11. Tsvetkov V.Ya. Informatsionnye edinitsey soobshchenii // Fundamental'nye issledovaniya. 2007, №12. s. 123-124.
12. Huber, D. (1989, May). A statistical approach to the segmentation and broad classification of continuous speech into phrase-sized information units. In Acoustics, Speech, and Signal Processing, 1989. ICASSP-89. 1989 International Conference on (pp. 600-603). IEEE.
13. Kratzer A. Facts: Particulars or information units? //Linguistics and philosophy. 2002. T. 25. № 5. S. 655-670.

14. Tajima K. et al. Discovery and Retrieval of Logical Information Units in Web // WOWS. 1999. S. 13-23.
15. Tsvetkov V.Ya. Logic units of information systems // European Journal of Natural History. 2009. № 2. p. 99-100.
16. Nandi A., Jagadish H.V. Qunits: queried units in database search // arXiv preprint arXiv: 0909.1765. 2009.
17. Tsvetkov V.Ya. Information Units as the Elements of Complex Models // Nanotechnology Research and Practice, 2014, Vol.(1), № 1. R. 57-64.
18. Luciano Floridi In defence of the veridical nature of semantic information // European Journal of Analytic Philosophy. 2007. Vol. 3. No. 1. p. 31-41.
19. Tsvetkov V.Ya. Information Constructions // European Journal of Technology and Design, 2014, Vol.(5), № 3. P. 147-152.
20. Barker K.J. et al. Information collection architecture and method for a data communications network : pat. 5375070 SShA. 1994.

УДК 004.8 004.9

Систематика информационных единиц

Татьяна Алексеевна Ожерельева

Московский государственный университет геодезии и картографии,
Российская Федерация
Старший преподаватель
E-mail: ozerotana@yandex.ru

Аннотация. Статья анализирует состояние и информационных единиц в науках об информации. Статья показывает многоаспектность применения информационных единиц. Рассмотрены примеры информационных единиц. Даны разные критерии типизации информационных единиц по группам. Дано деление информационных единиц на группы по применению и по отражению объективной реальности. Описаны проблемы теории информационных единиц.

Ключевые слова: информация; информационная модель; лингвистика; компьютерная лингвистика; информационные технологии; информационные единицы; информационные конструкции.