

ПРОБЛЕМЫ ПРИКЛАДНОЙ ЛИНГВИСТИКИ

Н. Г. Швец

ФОРМАЛЬНЫЙ СПОСОБ ПРЕДСТАВЛЕНИЯ ИЗОБРАЖЕНИЙ
РЕКЛАМНЫХ ОБЪЯВЛЕНИЙ В ВИДЕ ТЕЗАУРУСА

В статье рассматривается тезаурусное описание изображения как способ, позволяющий учесть как доминирующие, так и второстепенные черты изображения, а также представить иллюстрацию в виде некоторого набора слов (дескрипторов) и отношений между этими словами в рамках определенной предметной области.

Детально исследуется методика тезаурусного описания изображений при решении задач выбора компьютером иллюстрации к тексту рекламного объявления и порождения текста рекламного объявления по изображению.

Развитие семиотики дает возможность рассмотреть такие знаковые системы, как естественный язык и изобразительное искусство с некоторой единой точки зрения. Как и в естественном языке, в «языке живописи» есть свой набор основных единиц (словарь) и определенные правила их сочетаемости (грамматика) [1, с. 103]. Если смысл текста раскрывается словами, то иллюстрация «говорит» на языке форм или других ее элементов (цветов (красок), размеров, образов и т.п.).

При создании компьютерной системы, моделирующей понимание текста и изображения рекламного объявления (РО) человеком, возникает проблема поиска способа эффективной формализации изображения. Необходимо придать иллюстрации «неизобразительный» вид, ибо соответствующие алгоритмы приспособлены для переработки лишь неких символических описаний. Под описанием изображения будем понимать более простое, чем само изображение, представление исходного изображения в некотором «неизобразительном» виде (набор чисел, граф, словесная форма).

Среди возможных способов формального представления иллюстрации особый интерес вызывает тезаурусное представление изображения в качестве способа, позволяющего учесть как доминирующие, так и второстепенные черты изображения. Данный метод позволяет описать иллюстрацию в виде определенного набора слов (дескрипторов) и отношений между этими словами в рамках некоторой предметной области.

Обычно *тезаурус* определяют как словарь, содержащий лексические единицы с явным указанием семантических отношений между ними [2, с. 148].

В работе С. Е. Никитиной [3] тезаурус рассматривается не только как средство информационного поиска, но и как инструмент теоретических исследований в области семантики, гносеологии, лингвистики и терминоведения. Эти идеи получили развитие в другой работе того же автора [4].

Исследователь Е. А. Шингарева предлагает семиотическую модель тезауруса, в которой под тезаурусом в плане содержания понимается множество понятий и отношений между ними. В плане выражения – множество терминов, упорядоченных относительно друг друга. Денотатом тезауруса является представление (коллективный мыслительный образ) об объектном ядре предметной области. В семиотическую модель тезауруса введен также уровень референтов, «воплощающих те реальные объекты внешнего мира, которые на уровне представления о них включаются в процесс семиозиса» [5, с. 3].

Тезаурус может задавать парадигматические отношения: синонимии, родо-видовые, ассоциативные, синтагматические: объект – характерное свойство и др. При построении тезаурусов выявляют наиболее существенные отношения между словами, обусловленные исключительно свойствами самих предметов и явлений, обозначаемых этими словами, или, иначе говоря, содержанием выражаемых ими понятий. Из числа этих связей фиксируют лишь те, которые отражают существенные зависимости и типичны для рассматриваемой предметной области.

Лексические единицы тезауруса принято делить на две группы:

- 1) *термины* (ключевые слова, дескрипторы и синонимы);
- 2) *идентификаторы*, объединяющие имена собственные.

По определению, данному Т. Р. Кияк, «термин – слово или словесный комплекс, соотносящийся с понятием определенной организованной области познаний (науки, техники), вступающий в системные отношения с другими словами и словесными комплексами и образующий вместе с ними в любом отдельном случае и в определенное время замкнутую систему, отличающуюся высокой информативностью, однозначностью, точностью и экспрессивной нейтральностью» [6, с. 104].

В роли дескрипторов, по сложившейся практике построения тезаурусов, выступают термины, обозначающие некоторые понятия предметной области и удовлетворяющие принципам общеупотребительности, распространенности, краткости и терминологической точности.

Синонимами дескрипторов считаются термины, обозначающие то же понятие, что и дескриптор. В тезаурусе синонимами могут быть отдельные существительные, именные группы, отдельные прилагательные, глаголы и глагольные группы. Дескриптор в совокупности с его синонимами представляет в тексте определенное понятие предметной области.

Систематическая часть тезауруса представляет собой алфавитно упорядоченный перечень дескрипторов, или основных понятий с относящимися к ним лексическими единицами, и ключевых слов, являющихся разновидностью основных понятий и занимающих в систематической части свое место по алфавиту, но лишенных самостоятельной статьи. Как отмечает Ю. Н. Караулов, «ключевые слова являются теми нитями, которые связывают внутреннюю структуру тезауруса, т.е. организацию словарной статьи, с его внешней структурой, образуемой парадигматическими связями дескрипторов» [2, с. 150].

Стандартный тезаурус имеет, как правило, четыре входа, соответствующие четырем размерностям семантического представления лексики [7, с. 42]:

1. $K \rightarrow K$ – вход «от концепта к концепту», от одного понятия к другому, воплощенный в схеме иерархически и корреляционно связанных друг с другом понятий;

2. $K \rightarrow Z$ – вход, представленный дескрипторной частью тезауруса, отражающий переход «от концепта к знаку»;

3. $Z \rightarrow K$ – вход «от знака к концепту», соответствующий алфавитному указателю адресов слов в понятийных полях дескрипторов;

4. $Z \rightarrow Z$ – вход «от знака к знаку», или пермутационный указатель, который имеет смысл для тезаурусов, включающих словосочетания.

Процесс создания тезауруса состоит из следующих этапов [8, с. 19].

1. Выделение лексических единиц – формирование словника (гlossария) тезауруса.

2. Разработка набора семантических связей. Можно выбрать либо методику «от текста» (т.е. выявить те отношения, которые возникают между терминами в тексте), либо методику анализа предметной области, либо гибридный метод.

3. Актуализация связей – установление связей между терминами. При этом можно использовать знания экспертов, документы сферы фиксации (словари, классификаторы и т.д.), документы сферы функционирования (рефераты, статьи, монографии и т.д.).

В работе С. Д. Шелова сформулированы несколько общих требований, касающихся самого тезауруса и его построения [9, с. 6–7].

1. Соотношению «парадигматические связи в информационном тезаурусе – смысловые связи, которые задаются системой дефиниций соответствующих единиц» необходимо быть предельно ясным.

2. Структура тезауруса должна допускать математическую определенность, при которой каждая единица может быть охарактеризована своим уровнем.

3. Допуская произвольно глубокую дифференциацию парадигматических отношений, следует фиксировать в каком-то смысле только минимальное их количество.

4. Построение информационного тезауруса должно включать описание основных необходимых процедур анализа текста. Этим процедурам следует быть общими при решении многих задач автоматической обработки текста.

Анализ изображений 200 РО по теме «Косметика и парфюмерия» показал, что в большинстве случаев на иллюстрации обязательно присутствует изображение человека. Поэтому необходимо выявить наиболее употребительные существительные, используемые при описании внешнего облика человека и установить системные отношения между ними.

Отметим, что «особенности лексического состава описания определяются, прежде всего, структурой самого описания, которое служит для детального изображения состояний действительности с подробным перечи-

слением ряда одновременно существующих объектов и их признаков» [10, с. 96]. Функции объектов выполняют слова, обозначающие основные компоненты внешнего описания, которыми являются:

- 1) тело человека и его части (соматическая лексика);
- 2) мимика, поза, жесты (кинетическая лексика);
- 3) предметы одежды (вестиальная лексика).

Рассмотрим структурные отношения в группе соматической лексики. Для ее систематизации подходит отношение меронимии. Меронимия – способ ассоциативной сетевой организации понятий на основе отношения ЦЕЛОЕ – ЧАСТЬ. Это отношение обладает следующими свойствами [11, с. 30]:

- 1) оно транзитивно: ЦЕЛОЕ (ЦЕЛОЕ (X)) = ЦЕЛОЕ (X);
- 2) дескрипторы-части наследуют отношение АССОЦИАЦИЯ: АСЦ (ЦЕЛОЕ (X)) = АСЦ (X);
- 3) отношение ЧАСТЬ наследуется видовыми дескрипторами: ЧАСТЬ (ВЫШЕ (X)) = ЧАСТЬ (X);
- 4) отношение ВЫШЕ – НИЖЕ поглощается отношением ЦЕЛОЕ – ЧАСТЬ: ЦЕЛОЕ (ВЫШЕ (X)) = ЦЕЛОЕ (X), ВЫШЕ (ЦЕЛОЕ (X)) = ЦЕЛОЕ (X).

Пример построения меронимических отношений соматом представлен на рис. 1.

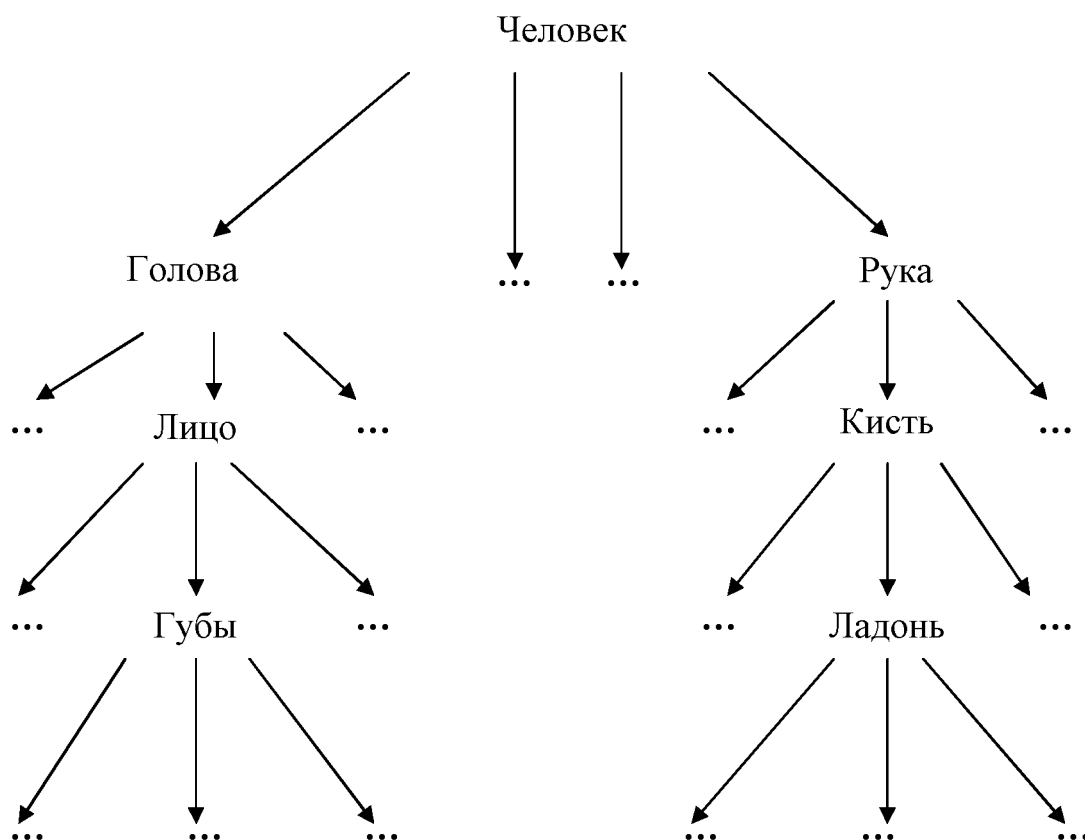


Рис. 1. Меронимические отношения соматом
(«→» – отношение меронимии; «...» – другие меронимы)

К характеризующим словам, описывающим внешние признаки, обычно относятся слова, указывающие на цвет, форму, толщину, величину. В зависимости от типа характеризуемого объекта набор признаков меняется. Анализ некоторых работ позволил нам выделить следующие признаки соматемы «Лицо» [10; 12]:

1) цвет и окраска кожного покрова (*загорелое, бледное, восковое, синее, землистое* и др.);

2) форма (*круглое, овальное, треугольное, квадратное* и др.);

3) величина (*большое, крупное, маленькое* и др.);

4) толщина (*худое, сухое, сытое, тонкое* и др.);

5) характер кожного покрова (*гладкое, холеное, морщинистое* и др.);

6) отличительная деталь (*длинноносое, курносое, скуластое, узколобое, большеглазое, черноглазое, чернобровое* и др.):

6.1) величина и форма носа;

6.2) форма скул;

6.3) величина и форма губ;

6.4) величина и форма лба;

6.5) величина, форма и цвет глаз;

6.6) наличие или отсутствие волосяного покрова на лице и голове.

7) состояние здоровья и возраст (*здоровое, усталое, молодое, свежее* и др.);

8) признак внешней выраженности интеллектуальных способностей (*умное, мудрое, простое, глупое* и др.);

9) общее впечатление:

9.1) положительные оценки (*красивое, миловидное, приятное* и др.);

9.2) отрицательные оценки (*неинтересное, невзрачное, непривлекательное* и др.);

10) доминирующая черта характера, особенность поведения (*благородное, честное, открытое, важное, самодовольное* и др.);

11) выражение черт характера и свойств личности (*волевое, властное, строгое, гордое, скромное, доброе, кроткое, хитрое, угодливое, насмешливое* и др.);

12) выражение доминирующих психоэмоциональных состояний (*живое, бодрое, радостное, счастливое, сердитое, равнодушное* и др.);

13) представление о социальном статусе личности (*благородное, интеллигентное* и др.).

Признаки других соматем представлены в таблице 1.

Признаки соматем

Соматема	Признак	Пример
Волосы	Цвет Форма Длина Густота Мягкость Толщина Влажность Блеск Сила Укладка Здоровье Тяжесть Чистота Общая положительная оценка Общая отрицательная оценка	<i>Русые</i> <i>Прямые</i> <i>Короткие</i> <i>Редкие</i> <i>Жесткие</i> <i>Тонкие</i> <i>Мокрые</i> <i>Блестящие</i> <i>Слабые</i> <i>Растрепанные</i> <i>Посеченные</i> <i>Легкие</i> <i>Грязные</i> <i>Восхитительные</i> <i>Некрасивые</i>
Глаза	Цвет Форма Величина Блеск Возраст Здоровье Эмоции Доброта Общая положительная оценка Общая отрицательная оценка Образ из сферы флоры Образ из сферы фауны Образ на основе экзотизмов	<i>Голубые</i> <i>Круглые</i> <i>Маленькие</i> <i>Блестящие</i> <i>Молодые</i> <i>Воспаленные</i> <i>Восторженные</i> <i>Злые</i> <i>Притягательные</i> <i>Невыразительные</i> <i>Ореховые</i> <i>Змеиные</i> <i>Цыганские</i>
Губы	Цвет Форма Толщина Величина Состояние Влажность	<i>Красные</i> <i>Оттопыренные</i> <i>Пухлые</i> <i>Большие</i> <i>Свежие</i> <i>Сухие</i>
Зубы	Цвет Форма Величина Расположение Состояние	<i>Белые</i> <i>Ровные</i> <i>Мелкие</i> <i>Редкие</i> <i>Жемчужные</i>

Соматема	Признак	Пример
Кожа	Цвет Толщина Пигментация Возраст Особое состояние Влажность Тип Эластичность	<i>Белая Тонкая Смуглая Молодая Гусиная Сухая Комбинированная Дряблая</i>
Нос	Форма Величина Цвет Характер кожного покрова Длина Образ на основе сходства по форме с носами животных и птиц	<i>Курносый Большой Красный Блестящий Короткий Орлиный</i>
Ресницы	Цвет Величина Длина Густота	<i>Черные Большие Длинные Редкие</i>
Тело (фигура)	Величина Толщина Тяжесть Сложение Гибкость Длина Нагота Цвет Чистота Здоровье Возраст Сила	<i>Щуплое Худое Легкое Стройное Гибкое Короткое Полуобнаженное Розовое Грязное Больное Молодое Дюжее</i>
Женщина	Возраст Осанка Рост Телосложение Сила Нагота Чистота Здоровье Ухоженность Интеллект Психозомоциональное состояние Характер Социальный статус Общая положительная оценка Общая отрицательная оценка	<i>Молодая Грациозная Высокая Стройная Сильная Голая Чистая Здоровая Ухоженная Умная Веселая Властная Интеллигентная Привлекательная Несимпатичная</i>

Очень часто на иллюстрации РО изображается предмет рекламы (флакон духов, баночка крема и др.). Для его описания используются такие признаки, как форма, цвет, содержимое, название, фирма-изготовитель, а для представления связей между дескрипторами – такие отношения, как ЦЕЛОЕ – ЧАСТЬ, АССОЦИАЦИЯ, ОБЪЕКТ – ХАРАКТЕРНОЕ СВОЙСТВО и др.

Можно предложить следующую методику тезаурусного описания изображений РО по теме «Косметика и парфюмерия», которая включает следующие этапы.

1. Выявление языкового, денотативного и коннотативного содержания для каждого изображения исследуемых РО.

2. Определение дескрипторов, образующих первый уровень тезауруса.

3. Выделение следующих типов отношений между дескрипторами: ЦЕЛОЕ – ЧАСТЬ, СОМАТЕМА – ПРИЗНАК, СОМАТЕМА – ДЕЙСТВИЕ, ОБЪЕКТ – ПРИЗНАК, ОБЪЕКТ – ДЕЙСТВИЕ, АССОЦИАЦИЯ.

4. Систематизация отношений типа СОМАТЕМА – ПРИЗНАК по ряду параметров (цвету, форме и т.п.). Такие параметры были выделены для описания следующих соматем: «Женщина», «Волосы», «Глаза», «Губы», «Зубы», «Кожа», «Лицо», «Нос», «Ресницы», «Тело (Фигура)».

5. Систематизация отношений типа ОБЪЕКТ – ПРИЗНАК по следующим параметрам: «форма», «цвет», «содержимое», «название», «фирма-изготовитель».

Проиллюстрируем процедуру тезаурусного описания на примере изображения РО, показанного на рис. 2.



Рис. 2. Рекламное объявление

Иллюстрация РО, представленного на рис. 2, имеет следующее языковое содержание, полученное путем рассмотрения надписи на бутылке с шампунем: «Новшество. Laboratoires Garnier Paris. Fructis с активным концентратом фруктов. Укрепляющий шампунь и бальзам-ополаскиватель. Подходит для частого применения. 2 в 1 для нормальных волос».

На анализируемом изображении можно выделить такие объекты, как рекламный персонаж (женщина), предмет рекламы (флакон с шампунем), фон (дольки фруктов), схема-пояснение (на ней показан корень волоса, окруженный молекулами разных активных веществ, которые его питают, укрепляют и разглаживают). Это денотативное содержание изображения.

Рассматривая исследуемое изображение более детально, можно выделить несколько коннотативных содержаний, т.е. тех ассоциаций, которые проходят через сознание человека при восприятии данного изображения.

Во-первых, завязывание волос в узел демонстрирует силу волос. Так как на иллюстрации изображена еще и бутылка с шампунем, это говорит о том, что именно он помог женщине так укрепить свои волосы.

Во-вторых, женщина и бутылочка с шампунем изображены на фоне долек с фруктами. Значит, активные вещества этих фруктов входят в состав рекламируемого шампуня.

В-третьих, желто-салатовые цвета фона и бутылки с шампунем ассоциируются с целебными свойствами предмета рекламы.

Следовательно, у человека, посмотревшего на данное изображение РО, возникает мысль: «Чтобы добиться такого эффекта сильных и блестящих волос, необходимо использовать шампунь, который рекламируется».

Тезаурусное представление анализируемого РО приведено в табл. 2.

Т а б л и ц а 2

Тезаурусное представление иллюстрации РО

Дескриптор (его уровень тезаурусной иерархии)	Мероним (его уровень тезаурусной иерархии)	Признак (его уровень тезаурусной иерархии)	Действие (его уровень тезаурусной иерархии) (Ассоциация)
Женщина (1)	Голова (2) Руки (2)		
Флаконт (1)	Шампунь (2)	Зеленый (2) Прямоугольный (2)	
Фон (1)	Фрукты (2)		
Схема- пояснение (1)	Волос (2) Вещество (2)		
Голова (2)	Волосы (3)		
Руки (2)			Завязывать волосы в узел (3) (сила волос)
Шампунь (2)		Fructis (Фруктис) (3) Garnier Paris (3)	
Волос (2)	Корень (3)		
Вещество (2)	Молекулы (3)	Активное (3) Зеленое (3)	
Фрукты (2)	Дольки (3)		
Волосы (3)		Блестящие (4) Сильные (4) Густые (4) Шелковистые (4) Длинные (4) Гладкие (4)	
Корень (3)		Здоровый (4) Сильный (4)	
Молекулы (3)		Зеленые (4) Красные (4)	Проникать в корень (4) (сила и блеск волос)
Дольки (3)		Зеленые (4) Желтые (4)	

Тезаурус является иерархической структурой, состоящей из дескрипторов различных уровней. Первый уровень тезауруса иллюстрации анализируемого РО (уровень тезауруса в табл. 2 обозначается цифрой в скобках) образуют следующие дескрипторы: *женщина*, *флакон*, *фон* и *схема-пояснение*.

Между дескриптором первого уровня *женщина* и дескрипторами второго уровня *голова* и *руки* устанавливаются системные отношения типа ЦЕЛОЕ (холоним) – ЧАСТЬ (мероним).

Дескриптор первого уровня *флакон* связан с дескриптором второго уровня *шампунь* отношением ЦЕЛОЕ – ЧАСТЬ, а с дескрипторами второго уровня *зеленый* и *прямоугольный* – отношением ОБЪЕКТ – ПРИЗНАК.

Между дескриптором первого уровня *схема-пояснение* и дескрипторами второго уровня *волос* и *вещество* устанавливаются отношения типа ЦЕЛОЕ – ЧАСТЬ.

Третий уровень тезауруса образуют следующие дескрипторы: *волосы*, *завязывать волосы в узел*, *сила волос*, *Fructis (Фруктус)*, *Garnier Paris*, *корень*, *молекулы*, *активное*, *зеленое*, *дольки*, *волосы*.

Дескрипторами четвертого уровня тезауруса являются: *блестящие*, *сильные*, *густые*, *шелковистые*, *длинные*, *гладкие*, *здоровый*, *сильный*, *зеленые*, *красные*, *желтые*, *проникать в корень*, *сила и блеск волос*.

В результате использования тезаурусного подхода были построены многоуровневые тезаурусные описания изображений 200 РО по теме «Косметика и парфюмерия», которые вошли в базу данных компьютерной программы, позволившей проверить эффективность работы формальной модели, связывающей текст и иллюстрацию РО.

ЛИТЕРАТУРА

1. Иванов, В. В. Структурно-типологический подход к семантической интерпретации произведений изобразительного искусства в диахроническом аспекте / В. В. Иванов, В. Н. Топоров // Труды по знаковым системам: учен. зап. Тарт. гос. ун-та. – Тарту, 1977. – Вып. 411. – С. 103–119.
2. Караулов, Ю. Н. Лингвистическое конструирование и тезаурус литературного языка / Ю. Н. Караулов. – М. : Наука, 1981. – 366 с.
3. Никитина, С. Е. Тезаурус по теоретической и прикладной лингвистике / С. Е. Никитина. – М. : Наука, 1978. – 312 с.
4. Никитина, С. Е. Семантический анализ языка науки (на материале лингвистики) / С. Е. Никитина. – М. : Наука, 1987. – 276 с.
5. Шингарева, Е. А. О двух направлениях представления семантики текста (тезаурус и фрейм) / Е. А. Шингарева // НТИ. Сер. 2. – 1982. – № 8. – С. 1–8.
6. Кияк, Т. Р. Лингвистические аспекты терминоведения / Т. Р. Кияк. – Киев: УМК ВО, 1989. – 104 с.
7. Анализ метаязыка словаря с помощью ЭВМ / Ю. Н. Караулов [и др.]; под общ. ред. Ю. В. Караулова. – М. : Наука, 1982. – 96 с.
8. Тезаурус как средство описания систем знаний / П. И. Браславский [и др.] // НТИ. Сер. 2. – 1997. – № 11. – С. 16–22.

9. Шелов, С. Д. Об одном подходе к информационному тезаурусу / С. Д. Шелов // НТИ. Сер. 2. – 1982. – № 7. – С. 6–10.
10. Старикова, Г. В. Лексика портретных описаний / Г. В. Старикова. – Л., 1985. – 198 с.
11. Лукашевич, Н. В. Представление знаний в системе автоматической обработки текстов / Н. В. Лукашевич, А. Д. Салий // НТИ. Сер. 2. – 1997. – № 3. – С. 27–33.
12. Богуславский, В. М. Оценка внешности человека: словарь / В. М. Богуславский. – М. : ООО «Изд-во АСТ»; Харьков : Торсин, 2004. – 254 с.

The article discusses one of the approaches to the formal presentation of illustrations. Thesaurus is proposed as a way of description of the image, allowing to take into account both the dominant and minor features of the image. Details of the thesaurus image description methodology are outlined in solving problems of choosing a computer illustration to the text of the ad and the generation of an ad text based on the image.

Поступила в редакцию 17.01.2019