

ПРИНЦИПЫ ПОСТРОЕНИЯ СЕМАНТИЧЕСКИХ АННОТАЦИЙ СОДЕРЖАНИЯ ИЗОБРАЖЕНИЙ

PRINCIPLES OF ANNOTATING THE IMAGE CONTENT

Е.Г. Соколова (egsokolo@ccas.ru)

Вычислительный центр Российской академии наук им. А.А. Дородницына

М.В. Болдасов (boldasov@nm.ru)

Российский НИИ искусственного интеллекта

В статье обсуждаются принципы создания семантических аннотаций для описания содержания изображений сцен дискретных объектов, представленных фотографиями открытого пространства – городскими видами, пейзажами. Аннотации создаются на основе онтологии понятий данной предметной области в виде XML представлений. Аннотации описывают статическое содержание изображений, включая композицию изображения. Формализованные аннотации могут использоваться как для поиска информации, например, в библиотечных и музейных коллекциях изображений, так и для генерации естественно-языковых описаний изображений как общих, так и по запросу пользователя. В последние годы в интернет создаются информационные системы и каталоги для поиска информации в хранилищах, содержащих данные различной природы – тексты, изображения, звук. При этом поиск осуществляется по запросам на разных языках. Примером такой информационной системы является международная информационная сеть культурного наследия в Европе, созданная сообществом BRICKS (BRICKS, 2004). Система включает в себя поиск информации о музейных экспонатах, в том числе, о произведениях изобразительного искусства. Поиск информации об изображениях актуален также для музеев фотографий. Кроме поиска по библиографическим данным актуален поиск изображений по их визуальным параметрам и по содержанию изображения, т.е. по изображенным объектам.

Для описания содержания изображений в коллекциях используются три средства: текстовые аннотации на Естественном Языке (ЕЯ), ключевые слова и перечень объектов.

Текстовые аннотации имеют три недостатка: а) использование конкретного ЕЯ затрудняет межъязыковой поиск; б) выражения на ЕЯ часто неоднозначны; в) чтобы сравнить текст запроса на ЕЯ и поисковый образ требуются нетривиальные операции. Например, в (Emeljanov et al., 2000) предлагается искать изображения по текстовым аннотациям, причем преодолевать вариативность анализа методом проведения эквивалентных трансформаций текста. Такой метод предполагает решение двух сложнейших задач – создание модулей синтаксического анализа и семантической интерпретации предложений произвольного текста, а также аппарата перифразирования предложений произвольного текста. Эти задачи пока остаются на стадии исследования.

Набор ключевых слов – наиболее простая и универсальная форма описания содержания документов, включая изображения. Такая форма используется повсеместно, в том числе и в BRICKS. Она позволяет преодолеть многоязыковость, задав соответствия между дескрипторами на разных языках. Обычно коллекция имеет свою, иногда частично структурированную, схему индексации (Hollink et al., 2003). Однако ключевые слова задают только тематику документа, а не его содержание.

Перечень изображенных объектов используется в Государственной исторической библиотеке в Москве. Это более содержательное описание, чем ключевые слова, однако оно не позволяет искать группы объектов, находящихся в определенной связи друг с другом, например, двухэтажные загородные дома у реки, сидящий около стола человек в сюртуке и котелке и т.п. Расширение перечней объектов до связанных семантических аннотаций содержания изображений позволило бы организовать поиск связанных объектов в массиве семантических аннотаций.

В статье представлены результаты исследования принципов построения формализованных семантических аннотаций для описания содержания изображений. Средствами описания являются понятия, обозначающие типы объектов, отношения между объектами, а также свойства объектов и их значения. Понятия собраны в созданную авторами экспериментальную модель Онтологии понятий видимого мира.

Понятия онтологии являются элементами семантических аннотаций изображений, представленных в формализме XML. Материалом данного исследования послужили фотографии городской и сельской местности, природных ландшафтов, сделанные с земли или с небольшого возвышения – мост, гора, крыша дома и т.п., собранные разными людьми из разных источников – печать, интернет, домашний архив.

Семантика изображения

Изображения открытого пространства мы отнесем к классу изображений, который назовем «сцены дискретных объектов» (СДО). Этот класс характеризуется тем, что за изображением стоит структура изображенных объектов типа сети фреймов, организующая участвующие в нем объекты в некую рациональную сцену. Эта сцена может быть описана в терминах объектов и значений их свойств, а также пространственных отношений между объектами. Создавая семантическую аннотацию изображения мы пользуемся понятиями, которые используют люди, описывая изображение на ЕЯ. Т.е. обсуждаемые описания антропоцентричны

Опыт показывает, что разные люди, описывая картинку, обращают внимание на разные объекты, поэтому ЕЯ описания одной и той же фотографии разными людьми могут существенно различаться, т.е. описания картинок, составленные людьми, субъективны. Вот, например, два описания содержания фотографии, представленной на рис. 1:

(Описание 1) *На картинке показан летний день после грозы.*

На переднем плане – река. Судя по точке съемки, снимок сделан с лодки или с моста. Берега ровные, может быть, это не река, а канал. Гроза, скорее всего, или вот-вот начнется, или только что закончилась: в затянутом тучами небе видны просветы, трава подсвечена солнцем. На небе видна радуга.

Слева на берегу стоит большой трехэтажный дом с красной крышей.

(Описание 2) *На картине изображен деревенский пейзаж: река, ее два берега, серое небо. На левом берегу реки стоит небольшой белый домик. Других домов не видно, домик одиноко стоит посреди зеленых деревьев. Около левого берега стоит лодка. Сквозь серые облака пробивается разноцветная радуга.*

В Описании 1 дом назван «большим», видимо автор думал о загородных домах, для которых трехэтажный дом - большой. В Описании 2 этот же объект назван «небольшим домиком». Возможно автор горожанин и для него все загородные дома – небольшие. В Описании 2 среди изображенных объектов упоминается лодка у левого берега, в Описании 1 – нет. Оба описания, начав с представления картины, далее перечисляют объекты, начиная с реки, которая занимает центральное место на картине. Но дальше реализуются противоположные стратегии: в Описании 1 обсуждается небо и кончается домом, а в Описании 2 начинается с дома и кончается небом.



Рис. 1

Задача создания максимально объективного описания в виде семантической аннотации предполагает перечисление по возможности всех изображенных объектов, потенциально интересных пользователю с опорой на объективные данные разных видов, в частности:

- реальный *относительный* размер объекта на изображении, например, часть площади изображения, занимаемая объектом по отношению к площади всего изображения или другого объекта;
- перспективу - уменьшение размера объектов при удалении от наблюдателя;
- знания об устройстве, относительных размерах (например, в сравнении с размером человека) и функциональности объектов.

Различия в описаниях, приведенных в предыдущем примере, связаны с выбором разных стратегий изложения и с тем, что разные люди опирались при описании картинке на разные знания о ее предметной области (ПО). Поэтому можно считать, что описания картинке объективно относительно ее ПО. ПО «открытое пространство» мы представляем в Онтологии.

Онтология

В предлагаемой модели онтологии представлены понятия, используемые для описания объектов, составляющих изображения открытых пространств, с перечислением различаемых человеком типов и свойств объектов и отношений между ними. Понятия онтологии задаются ее статьями. Статья онтологии состоит из следующих разделов:

заголовок – сопоставление понятию онтологии слова-идентификатора, которое обозначает тип объекта на изображении (например, понятия *дом стена, человек, фотография*);

неформальное определение – текстовое описание на ЕЯ содержания данного понятия;

род – понятие данной онтологии, которое является родовым по отношению к данному;

вид – перечисление понятий данной онтологии, которые являются видовыми по отношению к данному;

значения атрибутов, которые фиксированы в статье онтологии для данного понятия,

онтологические отношения – представляют знания об устройстве и функциональности объектов, а также о типичном расположении объектов, (например, *часть, место*);

устойчивые визуальные свойства самого изображения объекта, например, образ, текстура, форма, цветовая гамма, пока указываются только иногда в виде неформального комментария. Увеличение интереса и попыток создания онтологии семантики изображения, например, (Hunter, 2001), (Beloozerov et al., 2005), в будущем позволяют надеяться на развитие исследований по распознаванию объектов на изображении.

Онтология содержит иерархию понятий открытого пространства для построения аннотаций содержания изображений. Ее статьи полностью независимы от языковой семантики, поскольку содержат понятия, представленные образами. В корне онтологии три типа понятий: - изображение, отношение, атрибут. Изображение – это самое общее понятие, которое включает и саму картинку и ее части и типы объектов реального мира, изображенные на картине. Ниже приводится несколько статей онтологии.

Отношения включают онтологические и пространственные отношения. Пространственные отношения (например, *справа, выше*), передают взаимное расположение объектов на изображении с точки зрения наблюдателя. Они используются в описаниях сцен – семантических аннотациях изображений.

Атрибуты описывают свойства и состояния объектов. Информация о состоянии может быть получена в результате анализа зрительного образа, например, *стоящий (человек), сидящий (кошка, птица)*. Таким образом, типы атрибутов расширены по сравнению с тем набором, который обычно используется в семантических исследованиях. Кроме «классических» свойств или атрибутов, например, **цвет, размер**, к атрибутам отнесены понятия, обычно выражаемые глаголами, например, **бегающий**. В Онтологии они представлены как значения атрибута типа объекта **поза**.

Корневым объектом онтологии является понятие «изображение»:

(ИЗОБРАЖЕНИЕ

(ОПРЕДЕЛЕНИЕ («Изображение как единство носителя и содержания»))

(РОД (*))

(ВИД (ИЗОБРАЖЕНИЕ-НОСИТЕЛЬ ИЗОБРАЖАЕМОЕ)))

(ИЗОБРАЖЕНИЕ-НОСИТЕЛЬ

(ОПРЕДЕЛЕНИЕ ("Изображение как совокупность пикселей"))

(РОД (ИЗОБРАЖЕНИЕ))

(ВИД (ФОТОГРАФИЯ КАРТИНА))

(ЧАСТЬ (ГРАНИЦА-ИЗОБРАЖЕНИЯ ЧАСТЬ-ИЗОБРАЖЕНИЯ (центр, левый верхний угол и др.) ЛИНИЯ ФРАГМЕНТ-ИЗОБРАЖЕНИЯ))

)

(ИЗОБРАЖАЕМОЕ

(ОПРЕДЕЛЕНИЕ ("Воспринимаемое зрением изображение на плоской поверхности - картине, фотографии, как содержание – конфигурация представленных объектов."))

(РОД (ИЗОБРАЖЕНИЕ))

(ВИД (ОТКРЫТОЕ-ПРОСТРАНСТВО (пейзаж)

ЗАМКНУТОЕ-ПРОСТРАНСТВО (бытовая сцена)

ИЗОБРАЖЕНИЕ-ПРЕДМЕТА (натюрморт)

ИЗОБРАЖЕНИЕ-ЧЕЛОВЕКА (портрет) СПЕЦИАЛЬНОЕ-ИЗОБРАЖЕНИЕ (медицинский, например, цитологический, препарат, фотография поверхности Земли из космоса)))

(ЧАСТЬ (ГЕО-ОБЪЕКТ ОБЪЕКТ))

)

(ГЕО-ОБЪЕКТ

(ОПРЕДЕЛЕНИЕ "объекты планетарного масштаба, которые онтологически присутствуют на любом изображении открытого пространства")

(РОД (ИЗОБРАЖАЕМОЕ))

(ВИД (ПОВЕРХНОСТЬ-ЗЕМЛИ НЕБО ГОРИЗОНТ))

)

(ПОВЕРХНОСТЬ-ЗЕМЛИ

(ОПРЕДЕЛЕНИЕ ("Часть открытого пространства, на которой располагаются все ОБЪЕКТЫ))

(РОД (ГЕО-ОБЪЕКТ))

(ВИД (СУША ВОДА))

(ЦЕЛОЕ (ОТКРЫТОЕ-ПРОСТРАНСТВО))

Классификация объектов включает все, что может находиться на изображении – *дома, водоемы, дороги, газоны, машины, растения, солнце*, и т.д. В описаниях содержания картин использовалась общеязыковая онтология WordNet (WordNet), например, (Hollink et al.). В нашей онтологии меньше понятий, чем в общеязыковых онтологиях, например, нет абстрактных понятий типа *идея, судьба, умный*, и т.п., но могут быть абстрактные атрибуты типа *возраст*, поскольку они связаны с внешним видом объекта на изображении. В ней также отсутствуют глаголы, значения которых в небольшой степени покрываются значениями атрибута *поза*. Эти ограничения естественны – они накладываются визуальным способом восприятия мира человеком в естественных условиях.

В WordNet уже содержатся многие понятия, которые необходимо включить в онтологию открытого пространства. Однако описывающие их статьи не могут быть просто скопированы. WordNet следует философской традиции (Vossen, 2003), т.е. умозрительного, а не визуального созерцания мира. Они содержат много лишней информации и не содержат некоторой нужной информации. Например, из семи значений существительного Earth (земля) в WordNet только два соответствуют нужным нам понятиям, это:

[S: (n) earth, ground (the loose soft material that makes up a large part of the land surface) "they dug into the earth outside the church"].

[S: (n) land, dry land, earth, ground, solid ground, terra firma (the solid part of the earth's surface) "the plane turned away from the sea and moved back over land"; "the earth shook for several minutes"; "he dropped the logs on the ground"]

Первое обозначает землю как материал, например, *куча земли*. Второе оказывается слишком размытым, в частности, приведенный в WordNet пример употребления:

“ has instance

S: (n) America (North America and South America and Central America) “

может относиться к ПО «съемки из космоса», но никак не с земли. Указанные непосредственные видовые понятия второго значения, включающие такие релевантные для ПО «открытое пространство, видимое с земли» понятия как:

S: (n) beachfront (a strip of land running along a beach) S: (n) coastal plain (a plain adjacent to a coast),

и опосредованно:

S: (n) forest, woodland, timberland, timber (land that is covered with trees and shrubs),

не включают многие релевантные понятия, например, *дорога, газон* и т.п. а также все водоемы (*вода*)– *река, море, бассейн* и т.п., которые тоже представляют видимую человеком поверхность Земли. Многие из этих понятий представлены в WordNet своим функциональным аспектом и, таким образом, не содержат указания на то, что это поверхность Земли, например, *дорога*:

S: (n) road, route (an open way (generally public) for travel or transportation)

S: (n) road (a way or means to achieve something) "the road to fame"

Наша онтология ближе к онтологиям, решающим задачи, связанные с представлением значений конкретного текста. К ним относятся, в частности, онтологии для машинного перевода, основанного на знаниях (Knowledge Based Machine Translation (КВМТ)), например, (Nirenburg, Raskin, 2004). Отличительной чертой таких онтологий является наличие синтагматических семантических отношений – тех, которые передают взаимосвязи объектов в описаниях. Это актантажные отношения типа Actor, Goal, отношения между объектами, например, Localisation и др. Наши описания представляют «до языковую» картину. Они не содержат предикатов и отношений с участниками, но содержат онтологические и пространственные отношения, например, *часть, место, выше, справа*.

В начальной версии Онтологии, составленной для 20 изображений, имеется до 20 атрибутов, 10 отношений и 100 объектов.

Описание сцены понятиями онтологии

В описании сцены понятиям онтологии сопоставляются их экземпляры – объекты, прототипы реальных участников сцены. Объекты связаны отношениями.

Использовать аппарат концептуальных (онтологических) отношений для описания содержания и, в частности, сцен, предложила Н.Н. Леонтьева (Леонтьева, 2006). Например, следующий набор из двух концептуальных отношений с заполненными слотами:

место (дорога, машина)

справа (машина, человек)

описывают сцену, которая может быть выражена на ЕЯ разными способами, например, «На дороге находится машина. Справа от нее находится человек», или «На дороге стоит человек. Слева от него - машина». Предикат «стоять» для описания данного представления могут быть получены на основе информации о позе человека.

Содержание изображения описывается в терминах понятий Онтологии. Для представления описания используется нотация XML. Структурная природа представлений XML позволяет описать декомпозицию изображения, аннотация в форме XML легко уточняется за счет добавления более мелких деталей изображения.

Аннотация изображения антропоцентрична, т.е. ориентирована на те понятия, которыми пользуется человек для описания изображений. Человек воспринимает изображение сразу и целиком и понимает, как структурировано его содержание. Описание на ЕЯ обычно начинается с общей характеристики изображения, например, в (Описании 1) – «летний день после грозы», в (Описании 1) – «деревенский пейзаж». Такая общая характеристика является результатом мгновенного обобщения и оценки сцены человеком. Однако автоматической системой она может быть дана только после подробного анализа семантической аннотации. Наши аннотации включают по возможности все изображенные объекты, в которых может быть заинтересован пользователь, их атрибутов и отношений.

Описывая изображение, человек объединяет изображенные объекты в группы в соответствии с понятием им композиции. Для описания композиции будем использовать те понятия, которые использует человек. Типичным способом отражения композиции является использование планов (layer) изображения. Планы – это группы объектов, находящиеся приблизительно на равных расстояниях от наблюдателя. Понятие планов используется практически во всех описаниях фотографий. Последний, самый удаленный от наблюдателя план является общим фоном (например, небо, стена или лес). Так, на фотографии, представленной на рис. 2 прозрачная надпись «Rosfoto Image collection» расположена на переднем плане, тротуар и на нем идущий человек находятся на втором плане, фоном является стена.



Рис.2

Планы незримо проводят горизонтальные линии на изображении, на которых располагаются группы приблизительно равноудаленных от наблюдателя объектов. Планы могут использоваться для оценки размеров объектов с учетом перспективы. Кроме планов для описания совокупностей объектов используются множественные объекты, такие как *лес*, и группы - объекты, подчиняющие себе в аннотации другие объекты, например, *группа людей (человек-1, человек-2, ... человек-n)*, *дом (крыша, окна, двери)*. Между корнем группы, например, *дом* и подчиненными ему объектами, например, *крыша*, могут существовать онтологические отношения, в данном случае, *часть*. Пространственные отношения могут устанавливаться с отдельными объектами группы и с группой целиком.

Объекты играют разную роль в описании содержания изображения. Большие объекты (обычно гео-объекты), вытянутые в направлении взгляда наблюдателя, например, дорога, река (см. рис. 1), могут участвовать в формировании композиции изображения. Обычно такой объект расположен близко к центру изображения и делит его на три части – сам объект, объекты справа и слева от него. На рис. 1 это река (или канал), левый берег и правый берег. Групповые объекты могут также организовываться по планам, например, в описании содержания изображения на рис. 1 левый берег можно разделить на три плана: берег с крутизной 20%, поросший травой, на нем дом в три этажа с красной крышей, дерево слева от него (1 план); деревья (2 план); берег, покрытый песком с крутизной 5%, деревья на нем и лодка на воде у берега (3 план). Небо с радугой относится к фоновому плану всего изображения.

Цель аннотирования - не создание абсолютно полного описания, а представление тех объектов и отношений между ними, которые могут заинтересовать пользователя. В том случае, если описание оказалось недостаточно подробным, аннотация может уточняться. Это не потребует полной переделки аннотации. Например, могут добавляться атрибуты и зависимые объекты в групповой объект, что не затрагивает остальную часть изображения. В описании могут быть введены числовые характеристики, например оценки размера объекта, крутизны берега или скалы, как это сделано выше. Ниже приведено XML примерное представление содержания фотографии, представленной на рис. 1:

```
<?xml version="1.0" encoding="windows-1251" ?>
- <picture id="picture1">
  <v-object concept="РЕКА" id="river-1" />
  <left-v-object concept="БЕРЕГ" id="shore-1">
    <layer id="layer-1">
      <property concept="ПОКРЫТИЕ" meaning="ТРАВА" />
```

```

- <property concept="КРУТИЗНА" meaning="20" />
- <object concept="ДОМ" id="house-1">
  <property concept="ЦВЕТ" meaning="белый" />
- <object concept="КРЫША" id="roof-1">
  <property concept="ЦВЕТ" meaning="КРАСНЫЙ" />
  </object>
- <object concept="РЯД-ОКОН" id="floo-1">
  <property concept="КОЛИЧЕСТВО" meaning="3" />
  </object>
<object concept="дерево" id="tree-1" />
</object>
</layer>
- <layer id="layer-2">
- <object concept="ДЕРЕВО" id="trees-1">
  <property concept="КОЛИЧЕСТВО" meaning="SOME" />
  </object>
</layer>
- <layer id="layer-3">
  <property concept="ПОКРЫТИЕ" meaning="ПЕСОК" />
  <property concept="КРУТИЗНА" meaning="5" />
- <object concept="ЛОДКА" id="boat-1">
  <property concept="ЦВЕТ" meaning="БЕЛЫЙ" />
  </object>
</layer>
</left-v-object>
- <right-v-object concept="БЕРЕГ" id="shore-2">
  <property concept="ПОКРЫТИЕ" meaning="ЗАРОСЛИ" />
  </right-v-object>
- <layer-last id="layer-last-1">
  <object concept="ЛЕС" id="forest-1" />
- <object concept="НЕБО" id="sky-1">
- <object concept="ОБЛАКО" id="cloud-1">
  <property concept="КОЛИЧЕСТВО" meaning="SOME" />
  </object>
<object concept="РАДУГА" id="rainbow-1" />
</object>
</layer-last>
- <relations>
  <relation type="слева" domain="house-1" range="tree-1" />
  <relation type="контакт" domain="shore-1" range="boat-1" />
</relations>
</picture>

```

Заключение

В статье сформулированы принципы создания семантических аннотаций содержания изображений. Показано, что следуя антропоцентрическому принципу можно создать семантическую аннотацию содержания изображения на основе онтологии понятий данной ПО. Средствами описания являются также группирование объектов, установление между ними пространственных отношений и использование атрибутов, представляющих, в частности, некоторые физические свойства объекта на изображении, например, крутизна (берега, скалы). Аннотация позволяет уточнять свойства и состав объектов, добавляя свойства и объекты – члены группы объектов или части объекта. Аннотации могут использоваться для информационного поиска и для поиска новой информации в массивах изображений¹, а также для автоматической генерации описаний изображений на ЕЯ. Онтология предусматривает возможность сопоставления понятиям устойчивых визуальных свойств изображенных объектов. Это позволило бы в последствии исследовать возможности распознавания объектов на изображениях и предложить формализацию правил для построения семантических аннотаций содержания изображений.

¹ Поиск новой информации в массивах данных – новое и активно развивающееся направление автоматической обработки документов, получившее название data mining.

Работа выполнена при частичной поддержке Российского гуманитарного научного фонда, грант 04-04-00185а и Российского фонда фундаментальных исследований, грант 04-07-90187в.

Авторы также выражают благодарность студентам четвертого курса РГГУ, принявшим участие в исследовании.

Список литературы:

1. *Леонтьева Н.Н.* Автоматическое понимание текста: системы, модели, ресурсы. М.: «Академия», 2006.
2. *Beloozerov N.V., Murashov D.M., Trusova Yu.O., Yanchenko D.A.* Searching for solutions in the image analysis and processing knowledge base // Pattern recognition and image analysis: Advances in mathematical theory and application. – MAIK “Nauka/Interperiodica”, 2005. – Vol. 15, No.2. –P. 361-364.
3. BRICKS Project. Building resources for Integrated Cultural Knowledge Services © 2004. // European Commission • Sixth Framework Programme (<http://www.brickscommunity.org/>)
4. *Emeljanov G.V., Krechetova T.V., Kurashova E.P.* Tree grammars in the problems of searching for images by their verbal descriptions // Pattern recognition and image analysis. – 2000. Vol. 10, N4. P.520-526.
5. *Hollink L., Shreiber G., Wielemaker J., Wielinga B.* Semantic annotation of image collections // S. Handschuh, M. Koivunen, R. Dieng, and S. Staab, editors, Knowledge Capture 2003. Proceedings Knowledge Markup and Semantic Annotation Workshop, Florida, USA, October 2003. P. 41-48. Также доступна в интетенет:
6. <http://www.cs.vu.nl/~guus/papers/Hollink03b.pdf>
7. *Hunter J.* Adding multimedia to the semantic Web – building an MPEG-7 ontology // Proceedings of International Semantic Web Working Symposium (SWWS), Stanford, July 30 - August 1, 2001
8. (<http://archive.dstc.edu.au/RDU/staff/jane-hunter/semweb/paper.html>) .
9. *Miller G.* WordNet: a lexical database for English. Communications of the ACM, 38(11):39-41. WordNet: (<http://wordnet.princeton.edu/>)
10. *Nirenburg, S., Raskin, V.* Ontological Semantics // Language, Speech, and Communication. - MIT Press, 2004.
11. *Vossen P.* Ontologies // The oxford handbook of computational linguistics. Ed. by R. Mitkov, Oxford univ. Press. 2003. p. 464-482.