

Постановка эксперимента по индексированию корпуса изображений с помощью специализированного тезауруса

Б.В.Добров, А.В.Сидоров, Т.Н.Юдина
НИВЦ МГУ;
Центр Информационных Исследований
{dobroff, alexs, yudina}@mail.cir.ru

Н.В.Лукашевич, А.Д.Салий
ИСКРАН;
Центр Информационных Исследований
louk@mail.cir.ru

О.А.Штернова
МГЛУ;
Центр Информационных Исследований

1 ВВЕДЕНИЕ

Прогресс в развитии сетевой инфраструктуры Интернет, удешевление компьютерной памяти и стремительный прогресс в области устройств цифровой фотографии делают задачу оперирования большими коллекциями изображений весьма актуальной.

Узким местом остается возможность быстрого ассоциирования с изображением качественного поискового образа.

Для некоторых специфических задач разработаны методы [1,2] автоматического анализа изображений на основе той или иной методики распознавания образов. В некоторых областях такой подход может являться эффективным, особенно для небольших хорошо структурированных коллекций, используя интерфейс нахождения наиболее похожего изображения.

Авторов данной работы интересует прежде всего большие (и очень большие) коллекции событийной фотографии - фотоархивы информационных агентств, в перспективе - личные коллекции пользователей цифровых фотокамер, оборудованных специальными устройствами описания изображений, например, звуковыми.

Абсолютное большинство реально функционирующих систем [3,4] поиска изображений хранят поисковый образ в виде списка ключевых слов. Подход является достаточно эффективным для узких предметных областей с ограниченным числом наблюдаемых объектов, например, для коллекции

снимков NASA (<http://images.jsc.nasa.gov/iams/html/pao/searchpao.htm>).

Анализ списков ключевых слов, приписанных в существующих системах [3] к изображениям, показывает, что имеет место большая проблема (наблюдается достаточно большое количество ошибок) адекватного описания содержания изображения набором ключевых слов. Действительно, любой эксперт, вынужденный оперировать искусственным языком, которым является язык словаря ключевых слов, неизбежно старается минимизировать набор используемых слов, что и ведет к неточностям в описании.

Другим подходом [5] является ассоциирование изображению текста, расположенного в окрестности места появления изображения. Такого рода индексирование легко организовать, однако, часто изображение [6], помещенное в текст, имеет условное отношение к содержанию текста, отражает только какую-нибудь одну характеристику содержания текста.

В настоящей работе исследуется эффективность сочетания естественно-языкового (ЕЯ) описания изображений типа событийной фотографии с возможностями автоматического индексирования получающихся описаний. При этом ориентация делается на моделирование пользователя (эксперта), "наговаривающего" в стиле "потока сознания" некоторое описание к снимку, сделанному цифровой фотокамерой.

Для такого ЕЯ описания изображения следует ожидать, что каждому изображению приписывается достаточно полное, но короткое описание. Поэтому без развитых функций обработки синонимии и иерархии понятий, только с использованием морфологических анализаторов, невозможно организовать информационную систему, обеспечивающую приемлемый уровень полноты поиска изображений.

© Вторая Всероссийская научная конференция
ЭЛЕКТРОННЫЕ БИБЛИОТЕКИ:
ПЕРСПЕКТИВНЫЕ МЕТОДЫ И ТЕХНОЛОГИИ,
ЭЛЕКТРОННЫЕ КОЛЛЕКЦИИ
26-28 сентября 2000г., Протвино

Ядром исследований является изучение и улучшение качества специального тезауруса, который содержал бы средства для адекватного расширения запроса пользователя при поиске изображений.

2 ТЕЗАУРУС ДЛЯ АВТОМАТИЧЕСКОГО ИНДЕКСИРОВАНИЯ

Известны тезаурусы для описания графических изображений, созданные в Библиотеке Конгресса США [10]. Однако этот терминологический ресурс, помогая людям-индексаторам точнее приписывать изображению соответствующий термин, не ориентирован на автоматическую обработку свободного текста.

Тезаурус Центра информационных исследований [7-9] представляет собой интегрированный терминологический ресурс, предназначенный для автоматической обработки текста, что проявляется в большей подробности связей между понятиями, возможности использовать глобальную иерархию связей. Тезаурус ЦИИ состоит из тезауруса по общественно-политической тематике (15000 понятий), тезауруса общего языка (18500 понятий), 5000 понятий – общие для тезауруса по общественно-политической тематике и тезауруса общего языка.

Для автоматического индексирования текстов к тезаурусу по общественно-политической тематике дополнительно подключаются подтезаурусы объектов, имеющих собственные имена - географические названия и терминология (6700 понятий), персонала, организация, исторические события (1200 понятий).

Путем модификации программного обеспечения сопровождения терминологического ресурса удалось совместить ведение интегрированного ресурса и задачу выделения тезауруса для индексирования изображений.

Во множество понятий, отнесенных экспертами к визуальному тезаурусу, вошли понятия, которые могут быть использованы для ЕЯ описания событийной фотографии, обозначающие:

- конкретные видимые объекты;
- характеристики изображения/части изображения как такового - цвета, размера, количества и т.п.;
- видимые действия, идентифицируемые по характерным отличительным деталям (=БЕЖАТЬ=, =АТАКОВАТЬ=);
- действия, идентифицируемые по косвенным признакам на основе общих энциклопедических знаний экспертов о характерных ситуациях (=ВЕРОВАТЬ=, =ПЕТЬ=).

Во множество понятий, не вошедших в состав визуального тезауруса, вошли понятия, обозначающие:

- детали восприятия органами чувств, иными чем глазами (запахи, звуки), при этом даже если общая ситуация может быть уловима по изображению (=ПЕТЬ=), но оттенки ситуации уже нераспознаваемы (=БАРИТОН=, =НЕПРИЯТНЫЙ ЗАПАХ= и т.п.);
- детали, не воспринимаемые зрением без соответствующих технических устройств (=ГАЛАКТИКА=, =ВИРУС=), хотя здесь следует

учитывать возможность наличия визуально идентифицируемых моделей данных объектов;

- ментальные процессы и абстракции (=МЫСЛЬ=, =ВСПОМНИТЬ=);
- различные аспекты общественных отношений
- различные аспекты общественных отношений (=ПРОИЗВОДСТВЕННОЕ ОБЪЕДИНЕНИЕ=, =БЮДЖЕТ=).

В целом оказалось, что в имеющемся терминологическом ресурсе не удастся выделить аналитические правила определения рамок визуально наблюдаемых объектов. Действительно, основное правило вывода в тезаурусе – наследование свойств вдоль иерархии транзитивных отношений (ВЫШЕ-НИЖЕ, ЦЕЛОЕ-ЧАСТЬ). Однако, для некоторых понятий, определяемых как визуально идентифицируемые их тезаурусно подчиненные объекты не являются таковыми (=МОЛОЧНЫЕ ПРОДУКТЫ= -- =СЛИВКИ=).

Из общего интегрированного ресурса экспертами были выделены понятия, которые можно визуально идентифицировать на изображениях типа событийной фотографии - всего более 17500 понятий. Из них 6500 понятий общественно-политической сферы, 3200 понятий общеязыковой сферы, и все понятия, относящиеся к объектам, имеющие собственное наименование.

3 ОПИСАНИЕ ЭКСПЕРИМЕНТА

После анализа доступных по Интернет ресурсов изображений были определены три коллекции изображений (подборка фотографий "Москва день за днем" с www.glasnet.ru/glasweb/gate/pictures/; фотографии журнала "Огонек" 1998-1999гг. с www.gopnet.ru/ogonyok/win/welcome.html; коллекция фотографий фотографа Евгения Козловского с www.agama.com/r_club/gallery/KOZL/), из которых сформирована коллекция из 2000 изображений для целей некоммерческого использования.

Коллекция изображений была разделена на девять частей (ABCD – 80 изображений, ABC, ABD, ACD, BCD – по 20, A, B, C, D – по 600) и распределена между четырьмя экспертами, так что каждый из экспертов должен был описать по 600 изображений, (например, A + ABC + ACD + ABD + ABCD = 600). Таким образом, 80 изображений (какие именно экспертам было неизвестно) описываются одновременно всеми экспертами, еще 80 описываются одновременно тремя экспертами.

На данном этапе экспертам принципиально не давались какие-либо инструкции по описанию изображения – следование какому-либо сценарию или структуре описания – изучалась модель свободного ЕЯ описания изображения неподготовленным человеком – массовым собственником изображений типа событийной фотографии.

Создана графическая оболочка для естественно-языкового описания изображений. Оболочка позволяет загрузить во внутреннюю базу данных файлы изображений в формате JPG (имеется опция пакетной

загрузки), приписать каждому изображению некоторое количество "заметок", содержащих описание изображения в жанре "потока сознания". Реализованы простейшие механизмы поиска, а также разграничение доступа - когда один эксперт не может видеть информации, приписанной другим экспертом.

4 НЕКОТОРЫЕ РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТА

Как и предполагалось, обычно текстовое описание является небольшим по объему – размер редко превышает 250 байт, при среднем размере 100-200 байт в зависимости от эксперта. При этом среднее время описания изображения колеблется около одной минуты, и за один сеанс эксперту было тяжело описать более 50-100 изображений.

Наибольший интерес представляет сопоставление описаний одного и того же изображения, выполненных разными экспертами, определение меры схожести описаний, а также метрик, в которых мера схожести является максимальной.

Данная постановка актуальна при решении следующей задачи:

- пусть в базе данных информационной системы, содержащей изображения, хранится одно из текстовых описаний;
- предположим, что имеется пользователь, который примерно представляет какого рода фотографию он хочет найти;
- описания, сделанными другими экспертами мы можем трактовать, как описание информационной потребности пользователя, осуществляющего поиск изображения;
- основным вопросом является оценка полноты (отношение количества релевантных документов в отклике системы на запрос к общему количеству релевантных документов) и точности (отношение количества релевантных документов в отклике к общему количеству документов в отклике) информационного поиска в такого рода системе, выработка рекомендаций для повышения значений этих основных характеристик.

Приведем два примера из множества ABCD описаний изображений.

Пример 1. Файл 19990422.JPG
(www.glasnet.ru/glasweb/gate/pictures/)

(Эксперт А)

Ярко-синее небо, ветки с маленькими листочками, верхняя часть храма белого цвета с зеленым куполом.

(Эксперт В)

На фоне темно-синего неба зеленые ветки деревьев, а за ними белая башня с зеленой верхушкой и эмблемой наверху.

(Эксперт С)

На фоне синего неба башня с зеленой верхушкой. Зеленые ветки деревьев.

(Эксперт D)

Башня с шатрового типа зеленым шпилем на фоне очень синего неба. Наверху что-то типа герба. Здание покрашено в розовый цвет с белой отделкой.

Пример 2. Файл 15-60-AB.JPG
(www.ropnet.ru/ogonyok/pict/199915/)

(Эксперт А)

Артист Филиппенко в пестром пиджаке и синей рубашке держит правой рукой кепку, смотрит внутрь черного портфеля. Сзади светлая стена.

(Эксперт В)

Мужчина с залысиной в голубой рубашке и клетчатом пиджаке наклонился над открытой сумкой и держит в руке шапку. Он улыбается.

(Эксперт С)

Рыжеволосый мужчина с залысиной в серо-черном пиджаке кладет в портфель черный круг.

(Эксперт D)

Актёр Александр Филиппенко засовывает в (вынимает из) портфеля гибкий диск грампластинок. Одет в голубую рубашку и серый с черной косой полоской пиджак. Портфель большой в толщину. Филиппенко лыс, с рыжими волосами по бокам головы, улыбается.

Как можно видеть, пересечение описаний по прямому морфологическому совпадению не превышает 25-30% (за вычетом стоп-слов, предлогов, оборотов). В первом примере - =небо= (4 эксперта из 4), =зеленый= (4 из 4), =башня= (3 из 4), =ветка= (3 из 4), =дерево= (2 из 4), =верхушка= (2 из 4). Во втором примере - =пиджак= (4 из 4), =рубашка= (3 из 4), =портфель= (3 из 4), =Филиппенко= (2 из 4), =мужчина= (2 из 4) и т.д.

Некоторое удивление у авторов вызвало достаточно большой разброс в идентификации цветов, причем не только для сложных оттенков, но и для основных цветов (=ярко-синий=, =темно синий=, =синий=, =очень синий=; =пестрый=, =клетчатый=, =серо-черный=, =серый=).

Несколько большее пересечение получается при учете широкой квази-синонимии:

- =актер=, =артист=;
- =рыжеволосый=, =рыжие волосы=;
- =Филиппенко=, =Александр Филиппенко=;

однако не намного поднимающей процент пересечения описаний (до 40%).

Мера сходства описаний может быть увеличена лишь при использовании иерархических отношений между понятиями (используем для иллюстрации реально существующие в тезаурусе связи):

- =ЗДАНИЕ= >> =ХРАМ=;
- =МУЖЧИНА= >> =АКТЕР=;
- =МУЖЧИНА= >> =ФИЛИППЕНКО=;
- =КРУГ= >> =ГРАМПЛАСТИНКА=;
- =ЭМБЛЕМА= >> =ГЕРБ=,

причем необходимо рассматривать как спуск, так и подъем по иерархиям.

Наибольший эффект может быть достигнут при значительно более широком использовании иерархии, когда сначала необходимо подняться вверх по иерархическим отношениям, а затем спуститься вниз (в эту схему укладывается и предыдущий случай).

Такая процедура увеличивает меру сходства описания изображения до 70-90% (тем самым и полноту поиска):

- =ШПИЛЬ= << =КРЫША= >> =КУПОЛ=;
- =ШАТЕР= << =КРЫША= >> =КУПОЛ=;
- =КЕПКА= << =ГОЛ. УБОР= >> =ШАПКА=.

Широкое расширение терминологического описания изображения может иметь существенные отрицательные стороны. Действительно, может значительно уменьшиться точность поиска.

Некоторый оптимизм здесь может быть связан с тем, что для объектов, не имеющих собственного наименования (абстрактных понятий), по нашим наблюдениям достаточно пользоваться только парой отношений «ВЫШЕ» - для подъема по иерархии и «НИЖЕ» - для спуска по параллельным ветвям. Для объектов, имеющих собственное наименование (конкретных понятий), необходимо использовать дополнительно отношения «ЦЕЛОЕ»-«ЧАСТЬ».

Негативный эффект уменьшения точности возможно удастся уменьшить за счет:

- организации ранжирования, в том числе при повышении ранга за счет:
- морфологического и терминологического совпадения;
- организации обратной связи – добавления вспомогательных подзапросов в результаты найденного.
- использования гибридных моделей поиска – с развитием иерархических рубрикаторов, описывающих детали, общие для всех видов событийной фотографии:
 - пространственно-временных характеристик изображения;
 - цветовую гамму изображения;
 - материально-объектную составляющую изображения.

5 СОПОСТАВЛЕНИЕ С ТЕЗАУРУСОМ ВИЗУАЛЬНЫХ ОБЪЕКТОВ

Текстовые описания были обработаны с помощью стандартной процедуры Автоматической Лингвистической Обработки Текстов (АЛОТ) Центра Информационных Исследований [12]. В качестве терминологического ресурса использовался подтезаурис визуальных объектов.

Приведем результаты обработки – терминологический образ текстового сообщения эксперта D для первого примера (терминологический образ строится на основе морфологической информации о тексте):

(Эксперт D)

Башня с шатрового типа зеленым шпилем на фоне очень синего неба. Наверху что-то типа герба. Здание покрашено в розовый цвет с белой отделкой.

Формальный терминологический индекс выглядит следующим образом:

M107725;125632.0 4
БАШНЯ (ВЫСОКОЕ И УЗКОЕ СООР...
БАШНЯ
M117310;125632.0 4
БАШНЯ (МНОГОЭТАЖНЫЙ ДОМ)
БАШНЯ
M133565;125632.0 4
БАШНЯ (ЗАЩИЩЕННОЕ СООРУЖЕНИЕ)
БАШНЯ
M113960;157574.9 17
ШАТЕР (ВИД КРЫШИ)
ШАТРОВЫЙ
M120543;157574.9 17
ШАТЕР (БОЛЬШАЯ ПАЛАТКА, ...
ШАТРОВЫЙ
T120138;187092.19 22
КЛАСС (ПОДРАЗДЕЛЕНИЕ В КЛАСС...
ТИП
M109478;138757.24 30
ЗЕЛЕНЬ ЦВЕТ
ЗЕЛЕНЬ
M119626;138757.24 30
ЗЕЛЕНЬ (БЛЕДНЫЙ (О ЦВЕТЕ ЛИЦА...
ЗЕЛЕНЬ
M119627;138757.24 30
НЕЗРЕЛЫЙ (НЕ СПЕЛЫЙ)
ЗЕЛЕНЬ
T115252;157819.32 37
ШПИЛЬ (СТЕРЖЕНЬ НА КРЫШЕ ЗДАН...
ШПИЛЬ
A106446;162512.42 45
ЦВЕТ ФОНА
ФОН
M109476;152802.53 58
СИНИЙ ЦВЕТ
СИНИЙ
M118604;152802.53 58
СИНИЙ (БЛЕДНЫЙ (О КОЖЕ)
СИНИЙ
M111119;143400.60 63
НЕБО (ВЕРХНЯЯ СТЕНКА РОТОВОЙ...
НЕБО
M112832;143400.60 63
НЕБО (ВИДИМОЕ НАД ЗЕМЛЕЙ ПРОСТ...
НЕБО
T120138;187092.83 86
КЛАСС (ПОДРАЗДЕЛЕНИЕ В КЛАСС...
ТИП
T2022;102403.88 92
ГЕРБ
ГЕРБ
T2526;103950.96 101
ЗДАНИЕ
ЗДАНИЕ

и т.д.

Здесь первая группа содержит ссылки на номер концепта в тезаурусе, номер соответствующего текстового входа, а также ссылку на местоположение в тексте. Символ «Т» обозначает однозначно устанавливаемое соответствие, «М» - один текстовый вход соответствует более чем одному концепту среди описанных в тезаурусе, «А» - текстовый вход встречается в тезаурусе ровно один раз, но помечен как многозначный. Сдвиг вправо означает, что программа индексирования не может принять решения на основе имеющейся информации, если программа может разрешить многозначность на основе иерархии связей тезауруса, то используются символы «Т_М» и «Т_А». Отметим, одно из значений термина «ЗЕЛЕНЫЙ» (как «НЕОПЫТНЫЙ») из описанных не выводится, так как не отнесен к визуально идентифицируемым понятиям.

Для улучшения качества терминологического индекса были приняты следующие решения – фильтруются такие термины как «ТИП» (=«КЛАСС») и т.п., для всех текстовых входов обозначающих цвета, фон и т.п. другие значения специально блокируются.

Тогда получаем:

M107725;125632.0 4
 БАШНЯ (ВЫСОКОЕ И УЗКОЕ СООР...
 БАШНЯ
 M117310;125632.0 4
 БАШНЯ (МНОГОЭТАЖНЫЙ ДОМ)
 БАШНЯ
 M133565;125632.0 4
 БАШНЯ (ЗАЩИЩЕННОЕ СООРУЖЕНИЕ)
 БАШНЯ
 M113960;157574.9 17
 ШАТЕР (ВИД КРЫШИ)
 ШАТРОВЫЙ
 M120543;157574.9 17
 ШАТЕР (БОЛЬШАЯ ПАЛАТКА, КРЫТАЯ...
 ШАТРОВЫЙ
 T109478;138757.24 30
 ЗЕЛЕНЫЙ ЦВЕТ
 ЗЕЛЕНЫЙ
 T115252;157819.32 37
 ШПИЛЬ (СТЕРЖЕНЬ НА КРЫШЕ ЗДА...
 ШПИЛЬ
 T106446;162512.42 45
 ЦВЕТ ФОНА
 ФОН
 T109476;152802.53 58
 СИНИЙ ЦВЕТ
 СИНИЙ
 M111119;143400.60 63
 НЕБО (ВЕРХНЯЯ СТЕНКА РОТОВОЙ...
 НЕБО
 M112832;143400.60 63
 НЕБО (ВИДИМОЕ НАД ЗЕМЛЕЙ ПРОСТ...
 НЕБО
 T2022;102403.88 92
 ГЕРБ
 ГЕРБ
 T2526;103950.96 101
 ЗДАНИЕ
 ЗДАНИЕ

T107458;147980.103 111
 КРАСИТЬ, ПОКРЫВАТЬ КРАСКОЙ
 ПОКРАСИТЬ
 T109480;151908.115 126
 РОЗОВЫЙ ЦВЕТ
 РОЗОВЫЙ ЦВЕТ
 T_M109494;133490.130 134
 БЕЛЫЙ ЦВЕТ
 БЕЛЫЙ
 T_M124458;145356.136 143
 ОТДЕЛКА (ТО, ЧЕМ ОТДЕЛЫВАЮТ,
 ОТДЕЛКА

Таким образом, тезаурус визуальных объектов обеспечивает достаточно высокую степень покрытия текстовых описаний.

Вместе с тем, короткий текст не позволяет эффективно разрешать многозначность терминов. Отметим, что для текстов нормативных актов и текстов сообщений СМИ с помощью тезауруса по общественно-политической тематике удастся правильно разрешать до 75% многозначных слов и словосочетаний, что позволяет игнорировать индекс по неподтвержденным многозначным терминам при загрузке в ИПС. В случае индексирования текстовых описаний изображений мы обязаны учитывать неподтвержденные многозначные термины, хотя и снабжая их меньшим весом.

6 ДАЛЬНЕЙШИЕ РАБОТЫ

Модифицирована доступная коллективу авторов Интернет-версия Межуниверситетской информационно-поисковой системы УИС "Россия" (www.cir.ru) [11, 12] для представления в виде результатов поиска не только текстовых документов, но и изображений, в том числе физически расположенных на других сайтах. Согласован формат загрузки индексирующей информации об изображении. В настоящее время корректируются процедуры расширения запроса по терминам (с подъемом вверх и спуском по параллельным ветвям), отлаживаются процедуры ранжирования с учетом наличия большого количества многозначных терминов в индексе, проводятся эксперименты с тестовой загрузкой данных.

Требуются некоторые дополнительные усилия для учета достаточно свободной лексики, используемой экспертами при описании изображений, прежде всего учета различных словообразовательных суффиксов.

Недавно ИТАР-ТАСС выставило свои архивы событийной фотографии в Интернет (www.phototass.ru), при этом изображения рубрицируются по-видимому вручную, причем каждое изображение сопровождается текстом короткого информационного сообщения. Планируется провести сопоставительный анализ состава рубрик, приписанных изображению, результатов автоматического индексирования по тезаурусу визуальных объектов.

7 ЗАКЛЮЧЕНИЕ

- 1) Сформирована тестовая коллекция изображений типа событийной фотографии в количестве 2000

штук, где каждое изображение снабжено от одного до четырех текстовыми описаниями, выполненными разными экспертами.

- 2) На основе большого тезауруса для автоматического индексирования текстов сформирован специальный подтезаурус (17500 понятий, более 31500 терминов) для индексирования содержания коротких текстовых фрагментов, приписанных к изображениям.
- 3) Большое количество терминов, которые могут быть использованы при ЕЯ индексировании событийных фотографий показывает, что использование только ключевых слов для описания изображений не может быть эффективно организовано.
- 4) Получены оценки сходства текстовых описаний, выполненных различными экспертами. Совпадение по морфологическому и терминологическому индексу (только с использованием синонимии) оказалось невелико и оценивается величиной 25-40%.
- 5) Полноту поиска можно обеспечить только с использованием достаточно широкого расширения тезаурусной окрестности запрашиваемых терминов (подъем к более общему понятию и затем подключение параллельных ветвей).
- 6) Получены оценки «снизу» трудоемкости, полноты и точности применения изложенного подхода для создания больших архивов событийной фотографии.

БЛАГОДАРНОСТИ

Данная работа выполняется частично при поддержке гранта Российского Фонда Фундаментальных Исследований № 98-07-91167э.

Список литературы

- [1] R. Manmatha, "Multimedia Information Retrieval at the Center for Intelligent Information Retrieval", Proceedings of the Symposium on Document Understanding Technology (SDIUT 99), 183-198 (cobar.cs.umass.edu/pubfiles/mm-28.ps)

- [2] LIM: Leiden Imaging and Multimedia Group (www.wi.leidenuniv.nl/home/lim/)

- [3] PhotoDisc (search.photodisc.com)

- [4] The Cline Library Image Database - Northern Arizona University (www.nau.edu/cline/speccoll/imagedb.html)

- [5] L.S. Sable, V. Hatzivassiloglou, "Text-Based Approaches for the Categorization of Images", Proceedings of the Third Annual Conference on Research and Advanced Technology for Digital Libraries, 1999. (www.cs.columbia.edu/~sable/research/ecdl99.ps)

- [6] Фотоновости ИТАР-ТАСС (www.infoart.ru/misc/photonew/index.htm); www.infoart.ru/company/photo/index.htm)

- [7] Н.В. Лукашевич, А.Д.Салий, "Представление знаний в системе автоматической обработки текстов", НТИ. Сер.2., 1997, N3.

- [8] Н.В. Лукашевич, "От общеполитического тезауруса к тезаурусу русского языка в контексте автоматической обработки больших массивов текстов", Труды международного семинара Диалог-99, т.2, 184-190.

- [9] N. Loukachevitch, A. Sali, B. Dobrov, "Thesaurus for Automatic Indexing: Structure, Development, Use", Proceedings of International Congress "Terminology and Knowledge Engineering", Vienna, Termnet, 1999, 343-355.

- [10] Thesauri for Graphic Materials - Library of Congress (<http://lcweb.loc.gov/lexico/>).

- [11] С.В. Журавлев, Т.Н.Юдина, "Информационная система "Россия", НТИ. Сер.2, 1995, N 3, 18-20.

- [12] Т.Н. Юдина, С.В. Журавлев, "Российский межуниверситетский ресурсный и аналитический центр по гуманитарным исследованиям", Вестник РФФИ, 1999, N 3 (специальный выпуск), "Наука и информационное общество" (193.233.79.157/pub/vestnik/V3_99/2_8.htm)