DL-Media: An Ontology Mediated Multimedia Information Retrieval System

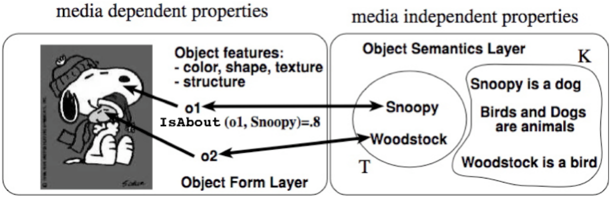
Umberto Straccia

ISTI-CNR, Pisa, Italy
straccia@isti.cnr.it

What is DLMedia?

- Multimedia Information Retrieval (MIR)
  - Retrieval of those multimedia objects of a collection that are relevant to a user information need
- DLMedia: is an ontology mediated MIR system, which combines
  - logic (semantic)-based retrieval
  - multimedia feature-based similarity retrieval
- An ontology layer is used to define (in terms of a description logic) the relevant abstract concepts
- A content-based multimedia retrieval system is used for feature-based retrieval

Illustrative, Conceptual Example (Logic-based MIR)



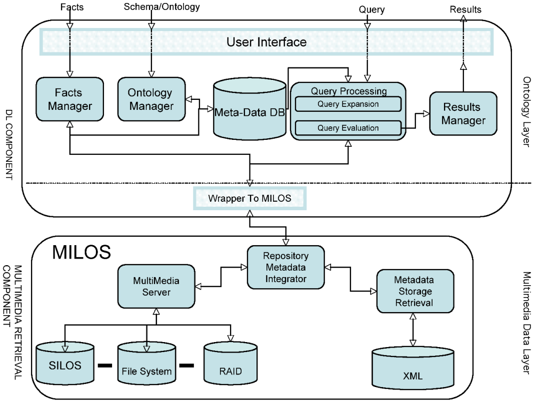
IsAbout		
ImageRegion	*Object ID*	*degree*
*o*1	*snoopy*	0.8
*o*2	*woodstock*	0.7
⋮	⋮	

"Find top-*k* image regions about white animals"

$$Query(x) \leftarrow ImageRegion(x) \wedge HasColor(x, white) \wedge isAbout(x, y) \wedge Animal(y)$$

The DL-MEDIA architecture

- From each multimedia object $o \in \mathbb{O}$ we automatically extract low-level features such as
  - text index term weights (object of type text)
  - colour distribution, shape, texture, spatial relationships (object of type image)
  - mosaiced video-frame sequences and time relationships (object of type video)
- All this pieces of data belong to the multimedia data layer
- On top of it we have the so-called ontology layer
  - defines the relevant concepts through which we may retrieve the multimedia objects $o \in \mathbb{O}$



- The ontology layer is managed by a Description Logic-based System
- The multimedia data layer is managed by the MILOS system

The Multimedia Retrieval Component

MILOS (Multimedia Content Management System),
http://milos.isti.cnr.it/

- General purpose multimedia software component supporting
  - multimedia data storage
  - content-based retrieval
  - multimedia metadata based on arbitrary XML metadata models
  - XML query language standards such as XPath and XQuery
- Is efficient and scalable w.r.t. storage and content-based retrieval

- Raw Data: text, images, video, audio
- Metadata: metadata about raw data
  - usually stored in XML format, e.g. MPEG7
- Query: keyword search, image similarity, ...
  - XQuery is a query language for querying XML data

MILOS Data Example

The funky lobby of the Blue Tree hotel in Brasilia


```
<SapirMMObject>
  <MediaLocator>
    <MediaUri>urn:milos:album:sapir:image_jpeg:b72e5db1fe28c8ebf0fa5fe2441f1d30</MediaUri>
  </MediaLocator>
  <photo id="1001779" secret="a217c7147f" server="1" farm="1" dateuploaded="1098481233" isfavorite="0" license="0" rotation="0" >
    <owner nsid="10249843@N00" username="klabrazil" realname="" location="" ></owner>
    <title>Blue Tree Brasilia</title>
    <description>The funky lobby of the Blue Tree hotel in Brasilia</description>
    <dates posted="1098481233" taken="2004-10-22 14:40:33" takengranularity="0" lastupdate="1102973777" ></dates>
    <comments>2</comments>
    <notes></notes>
    <tags>
      <tag id="90602-1001779-112696" author="10249843@N00" raw="bluetreebrasilia" machine_tag="0" >bluetreebrasilia</tag>
    </tags>
    <url>http://farm1.static.flickr.com/1/1001779_a217c7147f.jpg</url>
    <comments photo_id="1001779" >
      <comment id="90602-1001779-138311" author="10249843@N00" authorname="klabrazil" datecreate="1098482997" permalink="http://www.flickr.com/photos/10249843@
    </comment>
    </comments>
  </photo>
  <Mpeg7>
    <Description type="ContentEntityType" >
```

The DL-MEDIA architecture


```
<Mpeg7>
  <Description type="ContentEntityType" >
    <MultimediaContent type="ImageType" >
      <Image>
        <VisualDescriptor type="ScalableColorType" numOfBitplanesDiscarded="0" numOfCoeff="64" >
          <Coeff>85 48 63 5 1 -3 -5 15 -26 -22 6 -2 -4 4 -33 8 -5 -8 3 8 -9 -1 1 6 -8 6 -10 2 -8 5 1 -2 -3 -3 -3 -2 1 -3 -4 -1 -9 -7 -9 -1 -4 -6 0 -1 5 2 3 5
        </VisualDescriptor>
        <VisualDescriptor type="ColorStructureType" colorQuant="2" >
          <Values>32 0 0 0 0 0 0 11 32 16 0 0 0 0 0 0 12 23 0 5 0 0 13 6 52 30 18 12 35 21 2 2 29 70 140 126 27 27 27 27 103 42 25 22 83 43 24 22 132 155 15
        </VisualDescriptor>
        <VisualDescriptor type="ColorLayoutType" >
          <YDCCoeff>30</YDCCoeff>
          <CbDCCoeff>35</CbDCCoeff>
          <CrDCCoeff>30</CrDCCoeff>
          <YACCoeff5>26 28 23 12 13</YACCoeff5>
          <CbACCoeff2>15 18</CbACCoeff2>
          <CrACCoeff2>17 15</CrACCoeff2>
        </VisualDescriptor>
        <VisualDescriptor type="EdgeHistogramType" >
          <BinCounts>1 1 2 5 5 1 3 6 0 3 2 3 7 1 2 1 3 5 0 2 1 6 3 7 3 1 5 4 3 5 3 5 5 2 2 1 0 0 0 1 3 6 6 3 3 4 5 4 2 2 3 4 2 4 1 1 5 1 0 2 3 3 5 5 5 1 5 4
        </VisualDescriptor>
        <VisualDescriptor type="HomogeneousTextureType" >
          <Average>160</Average>
          <StandardDeviation>209</StandardDeviation>
          <Energy>242 237 221 208 218 246 230 188 189 181 199 200 206 177 157 169 166 179 169 146 142 144 162 155 175 179 138 117 151 170</Energy>
          <EnergyDeviation>246 241 217 207 219 247 219 179 181 179 189 190 191 171 142 155 149 169 160 121 137 126 155 155 167 160 109 102 114 151</EnergyDev
        </VisualDescriptor>
        <UsageInformation>
          <Rights>
            <RightsID>public</RightsID>
          </Rights>
        </UsageInformation>
        <CreationInformation>
          <Creation>
            <Title></Title>
            <Abstract>
```

- MILOS offers an advanced XML Search Engine (developed at ISTI-CNR)
  - Supports XQuery (with some limitations and extensions)
  - Offers image similarity search
  - Text search
  - Optimised for search intensive tasks
- XQuery:
  for \$a in /library//pictures
  where \$a/name = 'Brasilia'
  return \$a/location
- XQuery + Similarity:
  for \$a in /library//pictures
  where \$a/ColourDistribution $\approx$ '...'
  return \$a/location

MILOS has been tested within the following applications

- ECHO:
  - 50 hour of A/V data with IFLA-FRBR and MPEG-7 metadata (21 Gb of MPEG-1, 43.000 XML files)
- REUTERS:
  - 810.000 XML encoded, news agencies (2.6 Gb)
- DBLP and SIGMOD Records:
  - 187 Mb of XML files
- ANSA:
  - 1000 Color images, with MPEG-7 visual descriptor metadata
- PhotoBook:
  - On-line photo sharing: `http://milos.isti.cnr.it/milos/album` (more than 500 K of images)

MILOS similarity search is based on the metric space approach

- The similarity among two objects $o_1, o_2 \in \mathbb{O}$ is determined by a distance function

$$\begin{aligned} d : \mathbb{O} \times \mathbb{O} &\rightarrow [0, 1] \\ d(o, o) &= 0 \quad \text{(identity of indiscernibles)} \\ d(o_1, o_2) &= d(o_2, o_1) \quad \text{(symmetry)} \\ d(o_1, o_2) &\leq d(o_1, o') + d(o', o_2) \quad \text{(triangle inequality)} \end{aligned}$$

Supported similarity queries:

- Range Queries: given a query object $q \in \mathbb{O}$ and $r \in [0,1]$, find

$$Range(q,r) = \{o \in \mathbb{O} \mid d(q,o) \leq r\}$$

- *k*-Nearest Neighbors Queries: given a query object $q \in \mathbb{O}$ and natural number k, find

$$NN(q,k) = Top_k\{\langle o,s\rangle \mid o \in \mathbb{O}, s = d(q,o)\}$$

The Description Logic Component

- For computational reasons, DL-MEDIA is based on an variant of the DLR-Lite Description Logic
  - it is LOGSPACE w.r.t. the size of the data
  - but is NP w.r.t. the size of the ontology
- DLR-Lite is considered as a good compromise between expressive power and computational complexity, for data intensive applications

DL-MEDIA allows to specify the ontology by relying on axioms

- Consider *n*-ary relation symbols (denoted R) and unary relations, called *atomic concepts* (and denoted A)
- An axiom is of the form

$$Rl_1 \sqcap \ldots \sqcap Rl_m \sqsubseteq Rr \ ,$$

  where $m \geq 1$
  1. all Rl_i and Rr have the same arity
  2. where each Rl_i is a so-called *left-hand relation* and Rr is a *right-hand relation*
- Informally, read as "if Rl_1 and Rl_2 ... and Rl_m then Rr"

Examples (axioms involving atomic concepts)

- "Any italian city is an european city"

$$\text{ItalianCity} \sqsubseteq \text{EuropeanCity}$$

- "Any italian city, which is also big is a big european city"

$$\text{ItalianCity} \sqcap \text{BigCity} \sqsubseteq \text{BigEuropeanCity}$$

Examples (axioms involving *n*-ary relations)

- Assume we have a relation MyMetadata(docID, title, image, tag)
- We allow to make projection of the MyMetadata relation on some specified columns

$$\exists[1,3]\mathrm{MyMetadata} \sqsubseteq \exists[1,2]\mathrm{HasImageDescr}$$

$$\exists[1,4]\mathrm{MyMetadata} \sqsubseteq \exists[1,2]\mathrm{HasTag}$$

$$\exists[1,2]\mathrm{MyMetadata} \sqsubseteq \exists[1,2]\mathrm{HasTitle}$$

Examples (axioms involving *n*-ary relations)

- In case of a projection, we may further restrict it according to some conditions
- Assume we have a relation Person(firstname, lastname, age, email, sex)

$$\exists[2,3]\mathrm{Person} \sqsubseteq \exists[1,2]\mathrm{hasAge}$$

$$\exists[2,4]\mathrm{Person} \sqsubseteq \exists[1,2]\mathrm{hasEmail}$$

$$\exists[2,1,4]\mathrm{Person}.(([3] \geq 18) \sqcap ([5] =' \mathrm{male}')) \sqsubseteq \exists[1,2,3]\mathrm{AdultMalePerson}$$

Examples (axioms involving *n*-ary relations)

- We also allow to specify textual and image similarity conditions

$$(\exists[1]\mathrm{ImageDescr}.(([3]\ \mathrm{simImg}\ \mathrm{urn1}))) \sqcap (\exists[1]\mathrm{Tag}.(([2] =' \mathrm{sunrise}'))) \sqsubseteq \mathrm{Sunrise\_On\_Sea}$$

$$\exists[1]\mathrm{Title}.([2]\ \mathrm{simTxt}\ '\mathrm{lion}') \sqsubseteq \mathrm{Lion}$$

where *urn*1 identifies the image

Relation's Syntax

$$
\begin{array}{lll}
Rr & \longrightarrow & A \mid \exists[i_1, \ldots, i_k]R \\
Rl & \longrightarrow & A \mid \exists[i_1, \ldots, i_k]R \mid \\
 & & \exists[i_1, \ldots, i_k]R.(Cond_1 \sqcap \ldots \sqcap Cond_h) \\
Cond & \longrightarrow & ([i] \leq v) \mid ([i] < v) \mid ([i] \geq v) \mid ([i] > v) \mid \\
 & & ([i] = v) \mid ([i] \neq v) \mid \\
 & & ([i]\ simTxt\ 'k_1, \ldots, k_n') \mid ([i]\ simImg\ URN)
\end{array}
$$

DL-MEDIA Query Language

- A DL-MEDIA query consists of a conjunctive query of the form

$$q(\mathbf{x}) \leftarrow f(R_1(\mathbf{z}_1), \ldots, R_l(\mathbf{z}_l)) \ ,$$

x is a vector of variables, and every $\mathbf{z_i}$ is a vector of constants, or variables, *f* score combination function

$q(x) \leftarrow \text{Sunrise\_On\_Sea}(x)$
// find objects about a sunrise on the sea

$q(x) \leftarrow \text{CreatorName}(x, y) \wedge (y =' \text{paolo}') \wedge \text{Title}(x, z), (z \, \text{simTxt} \, '\text{tour}')$
// find images made by Paolo whose title is about 'tour'

$q(x) \leftarrow \text{ImageDescr}(x, y) \wedge (y \, \text{simImg} \, \text{urn2})$
// find images similar to a given image identified by *urn*2

$q(x) \leftarrow \text{ImageObject}(x) \wedge \text{isAbout}(x, y_1) \wedge \text{Car}(y_1) \wedge \text{isAbout}(x, y_2) \wedge \text{Racing}(y_2)$
// find image objects about cars racing

DL-MEDIA Semantics

To be compliant with the underlying MIR system MILOS, DL-MEDIA is based on mathematical fuzzy logic

- Given a concrete domain $\langle \Delta_{\mathrm{D}}, \Phi_{\mathrm{D}} \rangle$ with predicates on strings, numbers and images
- An *interpretation* $\mathcal{I} = \langle \Delta, \cdot^{\mathcal{I}} \rangle$ consists of
  - a *fixed infinite domain* Δ, containing Δ_{D}, and
  - an *interpretation function* $\cdot^{\mathcal{I}}$ that maps
    - every atom A to a function $A^{\mathcal{I}} \colon \Delta \to [0, 1]$
    - maps an n-ary predicate R to a function $R^{\mathcal{I}} \colon \Delta^n \to [0, 1]$
    - constants to elements of Δ such that $a^{\mathcal{I}} \neq b^{\mathcal{I}}$ if $a \neq b$ (unique name assumption).

DL-MEDIA Semantics (cont.)

- $\mathcal{I}$ *is a model of* (*satisfies*) an axiom $Rl_1 \sqcap \ldots \sqcap Rl_m \sqsubseteq Rr$ iff
$$\text{for all } \mathbf{c} \in \Delta^n, \min(Rl_1{}^{\mathcal{I}}(\mathbf{c}), \ldots, Rl_l{}^{\mathcal{I}}(\mathbf{c})) \leq Rr^{\mathcal{I}}(\mathbf{c}) \ ,$$
- $\mathcal{I}$ *is a model of* (*satisfies*) a query q the form $q(\mathbf{x}) \leftarrow \exists \mathbf{y} \phi(\mathbf{x}, \mathbf{y})$ iff for all $\mathbf{c} \in \Delta^n$:
$$q^{\mathcal{I}}(\mathbf{c}) \geq \sup_{\mathbf{c}' \in \Delta \times \cdots \times \Delta} \phi^{\mathcal{I}}(\mathbf{c}, \mathbf{c}')$$
- $\mathcal{I}$ *is a model of* (*satisfies*) $\langle q(\mathbf{c}), s \rangle$, iff $q^{\mathcal{I}}(\mathbf{c}) \geq s$
- $\mathcal{O}$ *entails* $q(\mathbf{c})$ to degree s iff each model $\mathcal{I}$ of $\mathcal{O}$ is a model of $\langle q(\mathbf{c}), s \rangle$
- The *greatest lower bound* of $q(\mathbf{c})$ relative to $\mathcal{O}$ is
$$glb(\mathcal{O}, q(\mathbf{c})) = \sup\{s \mid \mathcal{O} \models \langle q(\mathbf{c}), s \rangle\} \ .$$
- Basic inference problem: top-k retrieval problem
$$ans_k(\mathcal{O}, q) = \mathtt{Top}_k\{\langle \mathbf{c}, s \rangle \mid s = glb(\mathcal{O}, q(\mathbf{c}))\} \ .$$

Query Answering

Based on query rewriting of $q(\mathbf{x}) \leftarrow R_1(\mathbf{z}_1) \wedge \ldots \wedge R_l(\mathbf{z}_l)$

1. by considering $\mathcal{O}$, the user query q is *reformulated* into a set of conjunctive queries $r(q, \mathcal{O})$
   - For instance, given the query $q(x) \leftarrow A(x)$ and suppose that $\mathcal{O}$ contains the axioms $B_1 \sqsubseteq A$ and $B_2 \sqsubseteq A$, then we can reformulate the query into two queries $q(x) \leftarrow B_1(x)$ and $q(x) \leftarrow B_2(x)$, similarly as it happens for top-down resolution methods in logic programming
2. from the set of reformulated queries $r(q, \mathcal{O})$ we remove redundant queries
3. the reformulated queries $q' \in r(q, \mathcal{O})$ are translated to MILOS queries and evaluated. The query evaluation of each MILOS query returns the top-k answer set for that query
4. all the $n = |r(q, \mathcal{O})|$ top-k answer sets have to be merged into the unique top-k answer set $ans_k(\mathcal{O}, q)$. As $k \cdot n$ may be large, we apply the *Disjunctive Threshold Algorithm* (DTA) to merge all the answer sets

Preliminary Experiments

- 560.000 images together with their MPEG-7 metadata
  - The data has been provided by Flickr `http://www.flickr.com/`.
- 356 concept definitions
- 10 queries to be submitted to the system and measured for each of them
  - the precision at 10, *i.e.* the percentage of relevant images within the top-10 results
  - the number of queries generated after the reformulation process (q'_{ref})
  - the number of reformulated queries after redundancy elimination (q_{ref})
  - the time of the reformulation process (t_{ref})
  - the number of queries effectively submitted to MILOS (q_{MILOS})
  - the query answering time of MILOS for each submitted query (t_{MILOS})
  - the time of merging process using the DTA (t_{DTA})
  - the time needed to visualize the images in the user interface (t_{Img})
  - the total time from the submission of the initial query to the visualization of the final result (t_{tot})

Results:

Query	Precision	q'_{ref}	q_{ref}	t_{ref}	q_{MILOS}	t_{MILOS}	t_{DTA}	t_{Img}	t_{tot}
Q1	1.0	2	2	0.005	1	0.3	0	0.613	1.045
Q2	0.8	48	48	2.125	1	0.327	0	0.619	3.073
Q3	0.9	3	2	0.018	1	2.396	0	0.617	3.036
Q4	0.8	6	6	0.03	1	0.404	0	0.642	1.147
Q5	0.9	10	6	0.113	1	0.537	0	0.614	1.359
Q6	0.8	10	6	0.254	1	1.268	0	0.86	2.387
Q7	1.0	4	4	0.06	3	15.101	0.004	0.635	15.831
Q8	0.9	522	420	0.531	7	13.620	0.009	0.694	14.895
Q9	0.1	360	288	0.318	20	40.507	0.029	0.801	41.631
Q10	0.9	37	36	0.056	20	36.073	0.018	0.184	36.320

Conclusion & Outlook

- We've outlined the DL-MEDIA, *i.e.* an ontology mediated multimedia retrieval system
- Main features (so far) of DL-MEDIA:
  - DLR-Lite($\mathcal{D}$) like language as query and ontology representation language
  - supports feature-based queries, semantic-based queries and their combination
  - promisingly scalable
- A similar system has been developed that works on relational databases (postgres, mysql, ranksql)
  - DL-DB system: supports expressive top-k retrieval queries
  - Tested on Curricula Vitae matching (ca. 3000 OWL axioms, 10^5 records)
- Further investigating:
  - it seems reasonable to assume that the more specific the reformulated query becomes the less relevant may be its answers
  - multithreading of reformulated queries
  - allowing rules on top of axioms
  - to scale both to a DL-component with 10^4 concepts and to a MIR component indexing 10^6 images