What is DLMedia?

Umberto Straccia

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
What is DLMedia?

Illustrative, Conceptual Example (Logic-based MIR)

"Find top-k image regions about white animals"

```
Query(x) ← ImageRegion(x) ∧ HasColor(x, white) ∧ isAbout(x, y) ∧ Animal(y)
```
What is DLMedia?

The DL-MEDIA architecture

From each multimedia object $o \in \mathcal{O}$ 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 $o \in \mathcal{O}$
What is DLMedia?

The DL-MEDIA architecture

- 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
What is DLMedia?

The DL-MEDIA architecture

- **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
What is DLMedia?

The DL-MEDIA architecture

MILOS Data Example

The funky lobby of the Blue Tree hotel in Brasilia

<MediaLocator>
      <MediaUri>urn:milos:album:sopir:image_jpeg:b72e5db1fe28c8efb0fa5fe244f14d30</MediaUri>
</MediaLocator>
<photo id="1001779" secret="a217c7147f" server="1" form="1" dateuploaded="1098481233" isfavorite="0" license="0" rotation="0" />
      <owner nsid="10249843@N00" username="klabrazil" realname="" location="" />
      <title>Blue Tree Brasilia</title>
      <description>The funky lobby of the Blue Tree hotel in Brasilia</description>
      <dates posted="1098481233" token="2004-10-22 14:40:33" tagengranularity="0" lastupdate="1102973777" />
      <comments>2</comments>
      <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/10249843/1001779/"
      </comment>
      </comments>
</photo>

Umberto Straccia

DL-Media: An Ontology Mediated Multimedia Information Retrieval System
What is DLMedia?

The DL-MEDIA architecture
What is DLMedia?

The DL-MEDIA architecture

- 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

Umberto Straccia

DL-Media: An Ontology Mediated Multimedia Information Retrieval System
What is DLMedia?

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)
What is DL-Media?

The DL-Media architecture

MILOS similarity search is based on the **metric space approach**

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

\[
d : O \times O \to [0, 1]
\]

\[
d(o, o) = 0 \quad \text{(identity of indiscernibles)}
\]

\[
d(o_1, o_2) = d(o_2, o_1) \quad \text{(symmetry)}
\]

\[
d(o_1, o_2) \leq d(o_1, o') + d(o', o_2) \quad \text{(triangle inequality)}
\]
Supported similarity queries:

- **Range Queries**: given a query object $q \in \mathcal{O}$ and $r \in [0, 1]$, find
  \[
  \text{Range}(q, r) = \{ o \in \mathcal{O} \mid d(q, o) \leq r \}
  \]

- **$k$-Nearest Neighbors Queries**: given a query object $q \in \mathcal{O}$ and natural number $k$, find
  \[
  \text{NN}(q, k) = \text{Top}_k\{ \langle o, s \rangle \mid o \in \mathcal{O}, s = d(q, o) \}
  \]
What is DLMedia?

The DL-MEDIA architecture

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 $Rl_2$ \ldots and $Rl_m$ then $Rr$”
What is DLMedia?

The DL-MEDIA architecture

Examples (axioms involving atomic concepts)

- “Any italian city is an european city”
  
  ItalianCity ⊆ EuropeanCity

- “Any italian city, which is also big is a big european city”
  
  ItalianCity ∩ BigCity ⊆ BigEuropeanCity
Examples (axioms involving \(n\)-ary relations)

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

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

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

\[
\exists[1, 2]\text{MyMetadata} \sqsubseteq \exists[1, 2]\text{HasTitle}
\]
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)`

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

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

\[
\exists [2, 1, 4] \text{Person}.(([3] \geq 18) \sqcap ([5] = \text{'male'})) \sqsubseteq \exists [1, 2, 3] \text{AdultMalePerson}
\]
What is DLMedia?

The DL-MEDIA architecture

**Examples (axioms involving n-ary relations)**

We also allow to specify textual and image similarity conditions

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

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

where \text{urn1} identifies the image
What is DLMedia?

The DL-MEDIA architecture

Relation’s Syntax

\[ Rr \quad \rightarrow \quad A \mid \exists[i_1, \ldots, i_k]R \]

\[ Rl \quad \rightarrow \quad A \mid \exists[i_1, \ldots, i_k]R \mid \exists[i_1, \ldots, i_k]R.(\text{Cond}_1 \cap \ldots \cap \text{Cond}_h) \]

\[ \text{Cond} \quad \rightarrow \quad ([i] \leq v) \mid ([i] < v) \mid ([i] \geq v) \mid ([i] > v) \mid ([i] = v) \mid ([i] \neq v) \mid ([i] \sim \text{Txt}'k_1, \ldots, k'_n) \mid ([i] \sim \text{Img URN}) \]
A DL-MEDIA query consists of a conjunctive query of the form

\[ q(x) \leftarrow f(R_1(z_1), \ldots, R_l(z_l)), \]

\( x \) is a vector of variables, and every \( z_i \) is a vector of constants, or variables, \( f \) score combination function

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

\[ q(x) \leftarrow \text{CreatorName}(x, y) \land (y = '\text{paolo}' ) \land \text{Title}(x, z), (z \sim \text{Txt} '\text{tour}') \]
\[ // \text{find images made by Paolo whose title is about 'tour'} \]

\[ q(x) \leftarrow \text{ImageDescr}(x, y) \land (y \sim \text{Img} \text{urn2}) \]
\[ // \text{find images similar to a given image identified by urn2} \]

\[ q(x) \leftarrow \text{ImageObject}(x) \land \text{isAbout}(x, y_1) \land \text{Car}(y_1) \land \text{isAbout}(x, y_2) \land \text{Racing}(y_2) \]
\[ // \text{find image objects about cars racing} \]
What is DLMedia?

The DL-MEDIA architecture

DL-MEDIA Semantics

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

- Given a concrete domain $\langle \Delta_D, \Phi_D \rangle$ with predicates on strings, numbers and images
- An *interpretation* $\mathcal{I} = \langle \Delta, \cdot^\mathcal{I} \rangle$ consists of
  - a *fixed infinite domain* $\Delta$, containing $\Delta_D$, and
  - an *interpretation function* $\cdot^\mathcal{I}$ that maps
    - every atom $A$ to a function $A^\mathcal{I} : \Delta \rightarrow [0, 1]$
    - maps an $n$-ary predicate $R$ to a function $R^\mathcal{I} : \Delta^n \rightarrow [0, 1]$
    - constants to elements of $\Delta$ 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 \( Rl_1 \cap \ldots \cap Rl_m \subseteq Rr \) iff
  \[
  \text{for all } c \in \Delta^n, \min(Rl_1^\mathcal{I}(c), \ldots, Rl_m^\mathcal{I}(c)) \leq Rr^\mathcal{I}(c),
  \]

- \( \mathcal{I} \) is a model of (satisfies) a query \( q \) the form \( q(x) \leftarrow \exists y \phi(x, y) \) iff for all \( c \in \Delta^n: \)
  \[
  q^\mathcal{I}(c) \geq \sup_{c' \in \Delta \times \ldots \times \Delta} \phi^\mathcal{I}(c, c')
  \]

- \( \mathcal{I} \) is a model of (satisfies) \( \langle q(c), s \rangle \), iff \( q^\mathcal{I}(c) \geq s \)

- \( \mathcal{O} \) entails \( q(c) \) to degree \( s \) iff each model \( \mathcal{I} \) of \( \mathcal{O} \) is a model of \( \langle q(c), s \rangle \)

- The greatest lower bound of \( q(c) \) relative to \( \mathcal{O} \) is
  \[
  \text{glb}(\mathcal{O}, q(c)) = \sup\{s \mid \mathcal{O} \models \langle q(c), s \rangle\}
  \]

- Basic inference problem: top-\( k \) retrieval problem
  \[
  \text{ans}_k(\mathcal{O}, q) = \text{Top}_k\{\langle c, s \rangle \mid s = \text{glb}(\mathcal{O}, q(c))\}.
  \]
What is DLMedia?

Query Answering

Based on query rewriting of \( q(x) \leftarrow R_1(z_1) \land \ldots \land R_l(z_l) \)

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 \( \text{ans}_k(\mathcal{O}, q) \). As \( k \cdot n \) may be large, we apply the Disjunctive Threshold Algorithm (DTA) to merge all the answer sets
What is DLMedia?

The DL-MEDIA architecture

Preliminary Experiments

- 560,000 images together with their MPEG-7 metadata
  - The data has been provided by Flickr [http://www.flickr.com/](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 (\( t_{tot} \))
What is DLMedia?

The DL-MEDIA architecture

Results:

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

Umberto Straccia

DL-Media: An Ontology Mediated Multimedia Information Retrieval System
What is DLMedia?

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(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

Umberto Straccia

DL-Media: An Ontology Mediated Multimedia Information Retrieval System