Multimedia Retrieval and Reasoning

Umberto Straccia

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

1. A Basic Logic-based MIR Model
2. Logic-based Multimedia Annotation/Categorization
3. Logic-based Multimedia Retrieval
A Basic MIR Model

**Multimedia Information Retrieval (MIR)**
- Retrieval of those multimedia objects of a collection that are relevant to a user information need

**Logic-based MIR:**
- Ontology/Knowledge/Logic-based MIR, which combines
  - Logic (semantic)-based retrieval
  - Multimedia feature-based similarity retrieval
- Application domain knowledge is used to interpret multimedia object’s semantics
The use of logic for MIR is pretty old

  - Based on the estimation of the degree of implication between document $d$ and query $q$, $d \rightarrow q$


Some good sources to previous work (most ideas have been addressed already):


- FERMI. Formalisation and Experimentation on the Retrieval of Multimedia Information. ESPRIT Basic Research Action, no. 8134, 1994-1997 (http://www.dcs.gla.ac.uk/fermi/).
Any MIR model always starts from the identification of a suitable retrieval model [Str99], i.e. of a formal specification of the three basic entities of retrieval:

- the representation $d$ of multimedia documents $D$;
- the representation $q$ (called query) of users’ information needs $Q$; and
- the retrieval function $\mathcal{R}$, assigning a set of documents $d$ to each information need $q$.

To each retrieved document $d$ w.r.t. a query $q$ a degree of (system) relevance is given, called retrieval status value $RSV(d, q)$ indicating the confidence the system has in being a document $d$ relevant to the query $q$. 

Umberto Straccia
Multimedia Retrieval and Reasoning
A retrieval model can be characterized formally as follows:

- let $L_{Doc}$ be the language for representing multimedia documents $D$ as $d \in L_{Doc}$;
- let $L_{Query}$ be the language for representing users’ information needs $Q$ as $q \in L_{Query}$;
- let $dc \subseteq L_{Doc}$ be a collection of documents $d \in dc$.

A retrieval function $\mathcal{R}$ may be seen as a function

$$\mathcal{R}: 2^{L_{Doc}} \times L_{Query} \rightarrow 2^{(L_{Doc} \times [0,1])},$$

i.e. given a document collection $dc$ and a query $q$, $\mathcal{R}(dc, q)$ returns a set of pairs $\langle d, n \rangle$, where to each document $d$ the confidence, $n$, the system has in being the document $d$ relevant to the query $q$ is associated, i.e. $n = RSV(d, q)$.

Text Information Retrieval (vector space model): $L_{Doc} = L_{Query} = [0,1]^r$, $RSV(d, q) = \cos(d, q)$
Adequacy of a retrieval model depends on the choice of $L_{Doc}$, $L_{Query}$ and $R$

- Each of them plays an equally important role

MIR documents have (at least) two orthogonal dimensions: that of syntax, and that of semantics

**Syntax:**

*The syntax of a document is a collective name for all those features of the document that pertain to the medium that carries the document, and thus, are media dependent*

**Semantics:**

*The semantics of a document is a collective name for those features that pertain to the slice of the real world being represented, which exists independently of the existence of a document referring to it. Unlike form, the semantics of a document is thus media independent*
Corresponding to the two dimensions of a document just introduced, there are three categories of retrieval:

- **Syntax-based retrieval**
- **Semantics-based retrieval**
- **Combination of both of them**

**Syntax-based retrieval:**
Addresses, the syntactical (low-level features) properties of documents, e.g., word occurrences, color distributions, etc.

**Semantics-based retrieval:**
Rely on a symbolic representation of the meaning of documents, that is descriptions formulated in some suitable formal language.
MIR needs both dimensions to be taken into account, \textit{i.e.}

\[ L_{Doc} = \langle L_{Syntax}, L_{Semantics}, \text{Int} \rangle \]

\begin{itemize}
  \item \( L_{Syntax} \): language for representing each "object" (or "part") of interest in a document;
  \item \( L_{Semantics} \): language for describing the semantics of these objects;
  \item \( \text{Int} : L_{Syntax} \rightarrow L_{Semantics} \): mapping which associates a meaning description to each "object" of interest at the form level
\end{itemize}

\textbf{Interpretation function}

\begin{itemize}
  \item bridge between the syntax dimension and the semantics dimension determining the semantical meaning of the relevant objects identified at the syntax level
\end{itemize}

\textbf{Logic-based MIR}

\begin{itemize}
  \item \( L_{Semantics} \) and \( L_{Query} \) are logics
  \item Integration between \( L_{Syntax} \) and \( L_{Semantics} \) is defined in logical terms
  \item \( \mathcal{R} \) can be defined in terms of logical entailment
\end{itemize}
Example

**SYNTAX**

**SEMANTICS**

Background knowledge:
- "Woodstock is a bird."
- "Snoopy is a dog."
- "Birds are animals."
- "Dogs are animals."

Features:
- Color, Shape, Texture
- Structure

Interpretation

"Woodstock"

"Snoopy"

"Snoopy and Woodstock, sweetly embracing"
Interpretation functions are, in general,
- subjective
- imprecise

Subjectivity: what is this image about? A house? A bush?
Imprecision: is it about a house? To some extent . . .

\[ \text{Int} : L_{Syntax} \rightarrow (L_{Semantics}, [0, 1]) \]

Automatic interpretation (annotation/classification) is currently the most difficult task
Basic MIR Model Ingredients

\[ L_{\text{Doc}} = \langle L_{\text{Syntax}}, L_{\text{Semantics}}, \text{Int} \rangle \]

- **L_{\text{Syntax}}** ingredients:
  - Atomic regions
  - Region: atomic region or aggregation of regions

\[ \text{Region} = [A_1:T_1, \ldots, A_n:T_n] \]

where \( A_1, \ldots, A_n \) are the attributes (of type \( T_1, \ldots, T_n \)) of a region

- \( \text{Int} \) : Region \( \times \) Individual \( \rightarrow [0, 1] \)
  - E.g., \( \text{Int}(o_1, \text{snoopy}) = 0.8 \) means: “region \( o_1 \) is about \text{snoopy} to degree 0.8”

- \( L_{\text{Semantics}} \): a logic + Concrete domains supporting access to \( L_{\text{Syntax}} \)
\( L_{\text{Query}} \): e.g., Conjunctive queries supporting all three types of queries

Retrieval Function \( \mathcal{R} \): based on logical entailment

Example: “Find top-\( k \) image regions about white animals”

\[
\text{Query}(x, \text{score}) \leftarrow \text{ImageRegion}(x), \text{HasColor}(x, \text{white}, \text{score}_1), \text{Int}(x, y, \text{score}_2), \text{Animal}(y), \text{score} = \text{score}_1 \cdot \text{score}_2
\]

\( \text{HasColor}(x, \text{white}, \text{score}_1) \): syntax-based query part (\( \text{score}_1 \) = degree of being region \( x \) white)

\( \text{Int}(x, y, \text{score}_2), \text{Animal}(y) \): semantics-based query part (\( \text{score}_2 \) = degree of being region \( x \) about \( y \))
Logic-based Multimedia Annotation/Categorization

- Annotation/Classification of Multimedia data corresponds to make explicit (to define) the region interpretation function

\[ Int : L_{Syntax} \times L_{Semantics} \rightarrow [0, 1] \]

- Most of the time, \( Int \) is specified manually with the help of a graphical user interface and/or region detector

- Sometimes, \( Int \) is learned automatically with the help of machine learning tool (with optionally a region detector)
Categorization in Machine Learning [Seb02] applied to multimedia:

Using a training set, learn $\Phi: L_{Syntax} \times L_{Semantics} \rightarrow [0, 1]$, that approximates $Int$

We speak of

- **Single Label Categorization**: A region is assigned to at most one category/individual only (e.g., assign $o$ to $c$ where $\Phi(o, c)$ got the highest score)
- **Multi-Label Categorization**: A region may be assigned to more than one category/individual (e.g., assign to $o$ the top-k ranked $c$, according to the score of $\Phi(o, c)$)
- Example: Whole image is region $o$ and a classifier computes

\[
\begin{align*}
\Phi(o, bush) & \mapsto 0.9 \\
\Phi(o, house) & \mapsto 0.8 \\
\Phi(o, grass) & \mapsto 0.7 \\
\Phi(o, tree) & \mapsto 0.6
\end{align*}
\]

Single Label Categorization: $Int(o, bush) = 0.9$, else $Int(o, c) = 0.0$
Multi-Label Categorization (k=2): $Int(o, bush) = 0.9$, $Int(o, house) = 0.8$ else $Int(o, c) = 0.0$

**Boolean Categorization**: $Int(o, c) \in \{0, 1\}$
Most image classification methods rely on $L_{Syntax}$ only

- Use of low-level features for classification

They typically rely on classical machine classification methods, such as

- Naive Bayes
- Support Vector Machines (SVM)
- k-Nearest Neighbors (kNN)

Some can be extended to rely on $L_{Semantics}$ and the combination of $L_{Syntax}$ and $L_{Semantics}$, as well
Example: kNN

- Given a training set \( TS = \{ o_1, \ldots, o_{|TS|} \} \) of image regions
- Given a set \( C = \{ c_1, \ldots, c_{|C|} \} \) of categories \( c_i \in L_{Semantics} \)
- Given an a-priori category assignment \( \tilde{Int}: TS \times C \rightarrow \{0, 1\} \)
  - \( \tilde{Int}(o, c) = 1 \) means that training image region \( o \in TS \) has been (manually) assigned to category \( c \)
- Given \( \text{Top}_k: \text{Region} \rightarrow TS \), where \( \text{Top}_k(o) \) is the set of \( k \) regions \( o_i \in TS \) that maximize \( RSV(o, o_i) \)
- Category Status Value (CSV) of \( o \in \text{Region} \) w.r.t. category \( c \in C \):
  \[
  CSV(o, c) = \sum_{o' \in \text{Top}_k(o)} RSV(o, o') \cdot \tilde{Int}(o', c)
  \]

- Note: straightforward to implement over a MIR system
  - Submit query: “find top-\( k \) most similar images \( o' \in TC \) to \( o \)”
  - Go through the ranked list and compute \( CSV(o, c_i) \) for all \( c_i \in C \)
Notes

\[ CSV(o, c) = \sum_{o' \in \text{Top}_k(o)} RSV(o, o') \cdot \text{Int}(o', c) \]

- \( RSV(o, o') \) may be based on
  1. Visual low-level features in \( L_{\text{Syntax}} \) only (e.g., MPEG7 standard distance measures [Eid03]),
    \[ RSV(o, o') = RSV_{\text{Syntax}}(o, o') \]
  2. Semantics properties of \( L_{\text{Semantics}} \) only (use some semantic similarity function, e.g. [dFE06]),
    \[ RSV(o, o') = RSV_{\text{Semantics}}(o, o') \]
  3. The combination of both, e.g.
    \[ RSV(o, o') = \alpha \cdot RSV_{\text{Syntax}}(o, o') + (1 - \alpha) \cdot RSV_{\text{Semantics}}(o, o'), \quad \alpha \in [0, 1] \]

- In cases 2. and 3., logic-based reasoning comes into play, as usually semantic similarity functions rely on some logical reasoning
- We have a seamless integration of feature-and logic-based classification
\[ CSV(o, c) = \sum_{o' \in Top_k(o)} RSV(o, o') \cdot \tilde{Int}(o', c) \]

together with

\[ RSV(o, o') = \alpha \cdot RSV_{Syntax}(o, o') + (1 - \alpha) \cdot RSV_{Semantics}(o, o') \]

gives

\[ CSV(o, c) = \alpha \cdot CSV_{Syntax}(o, c) + (1 - \alpha) \cdot CSV_{Semantics}(o, c) \]
Sometimes, we have (or variants of)

\[ CSV(o, c) = \alpha \cdot CSV_{Syntax}(o, c) + (1 - \alpha) \cdot CSV_{Semantics}(o, c), \quad \alpha \in [0, 1] \]

- \( CSV(o, \text{sportcar}) = ? \)
- Suppose \( CSV_{Syntax}(o, \text{sportcar}) = 0.9 \)
- What about \( CSV_{Semantics}(o, \text{sportcar}) \)?
- Suppose \( KB \models Region(o) \land \exists y.\text{Int}(o, y) \land \text{AudiTT}(y) \land \ldots \)
- Is Audi TT a Sport Car? \( \text{Speed} = 243, \text{Accelleration} = 6.9 \)
- We can estimate from a training set (Naive Bayes)

\[
CSV_{Semantics}(o, \text{sportcar}) = \frac{Pr(\text{AudiTT}|\text{SportCar}) \cdot Pr(\text{SportCar}) \cdot (1 / Pr(\text{AudiTT}))}{Pr(\text{speed} \geq 243) \cdot Pr(\text{accel} \leq 6.9)} \\
\approx \frac{Pr(\text{speed} \geq 243|\text{SportCar}) \cdot Pr(\text{accel} \leq 6.9|\text{SportCar}) \cdot Pr(\text{SportCar})}{Pr(\text{speed} \geq 243) \cdot Pr(\text{accel} \leq 6.9)}
\]

\[
Pr(\text{speed} \geq 243|\text{SportCar}) = \frac{|\{ o \in TS | KB \models \exists y \exists z. \text{Region}(o) \land \text{Int}(o, y) \land \text{Sportcar}(y) \land \text{hasSpeed}(y, z) \land z \geq 243 \}|}{|TS|}
\]

...
Example cont.

- **Sport Car:**
  \[ \forall x, hp, sp, ac \; \text{SportCar}(x) \iff 0.3HP(x, hp) + 0.2Speed(x, sp) + 0.5Accel(x, ac) \]

- Each feature, gives a degree of truth depending on the value and the membership function

  \[
  \begin{align*}
  HP(x, hp) &= rs(180, 250)(hp) \\
  Speed(x, sp) &= rs(180, 240)(sp) \\
  Accel(x, ac) &= ls(6.0, 8.0)(ac)
  \end{align*}
  \]

- Degree of truth of **SportCar(AudiTT):**
  \[ 0.1 \cdot 0.28 + 0.3 \cdot 1.0 + 0.6 \cdot 0.55 = 0.658 \]
The fuzzy membership functions, as well as the weights, can be learned from a training set (large literature):

$$\begin{align*}
HP(x, hp) &= rs(192, 242)(hp) \\
Speed(x, sp) &= rs(193, 234)(sp) \\
Accel(x, ac) &= ls(6.5, 7.5)(ac)
\end{align*}$$

![Membership functions](image.png)

Learned Training Sport Class:

$$\forall x, hp, sp, ac \ \text{TrainingSportCar}(x) \iff 0.3HP(x, hp) + 0.2Speed(x, sp) + 0.5Accel(x, ac)$$

Now, a classification method can be applied: e.g. kNN classifier

$$\forall x, hp, sp, ac \ \text{SportCar}(x) \iff \sum_{y \in Top_k(x)} \text{Similar}(x, y) \cdot \text{TrainingSportCar}(y)$$

$$\forall x, hp, sp, ac \ \text{Similar}(x, y) \iff 0.3 \cdot HP(x, hpx) \cdot HP(y, hpy) + 0.2 \cdot Speed(x, spx) \cdot Speed(y, spy) + 0.5 \cdot Accel(x, acx) \cdot Accel(y, acy)$$

where $Top_k(x)$ is the set of top-$k$ ranked most similar cars to car $x$
Other methods: determine the logically consistent/most “reasonable” interpretations of a region $o$, given

- The aggregation of regions $o_j$ in which $o$ is involved
- The values of $CSV_{\text{Syntax}}(o, c_i)$ and $CSV_{\text{Syntax}}(o_j, c_k)$
- The background knowledge
- The semantic information about $o$ and the involved $o_j$

Many variants are possible

The effectiveness of such categorization methods has still to be shown

The efficiency/scalability of such methods has also to be shown (a scalable logic-based MIR system may be required)

In my humble opinion, I encourage to rely on well grounded categorization methods (due to their success) rather than on method based on the first item. This latter idea is very old (see, e.g. [BRT88]) and suffers on effectiveness and efficiency (yet)
We recall our basic logic-based MIR model

\[ L_{Doc} = \langle L_{Syntax}, L_{Semantics}, Int \rangle, \] where \( L_{Semantics} \) is a logic
\( L_{Query} \) is a logic
Retrieval function \( R \) function

\[ R : 2^{L_{Doc}} \times L_{Query} \rightarrow 2^{(L_{Doc} \times [0,1])}, \]

is defined on some notion of logical entailment, \( \models \)
“Find top-k image regions about white animals”

What do we choose concretely for all these ingredients?
Concerning $L_{Syntax}$. We may start from MPEG7 [Mul02] with some XML extension of it.

There are some systems supporting MIR over MPEG7 data,

For instance, 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
• 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
Which logic for $L_{semantics}$ and $L_{Query}$?
What semantics for $\models$ on which the retrieval function $\mathcal{R}$ is based?
A reasonable choice should be
- scalable in size
- $\models$ should accommodate the inherent uncertainty/vagueness of multimedia data interpretation and retrieval

Hence, we need a logic that
1. is tractable (at least in the size of multimedia data)
2. that supports the ranking of query results
3. supports top-k retrieval
4. seamlessly integrates with an underlying feature-based MIR system
Refer to [LS07, LS08, Str08, Str07]

- Logic that
  1. is **tractable** (at least in the size of multimedia data)
      - Datalog, DL-Lite, DLR-Lite, RDFS
  2. that supports the **ranking** of query results
      - Fuzzy/Probabilistic variants of Datalog, DL-Lite, DLR-Lite, RDFS
      - Note: Most of the time probabilistic variants increase complexity
  3. supports **top-k retrieval**
      - Algorithm known for Fuzzy variants of Datalog, DL-Lite, DLR-Lite
      - Unknown for fuzzy RDFS and probabilistic variants
  4. seamlessly **integrates** with an underlying feature-based MIR system
      - DL-**MEDIA** [SV08] is a system putting all together (but, see also HySpirit [FR97])
DL-MEDIA: 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.
The ontology layer is managed by a Description Logic-based System
The multimedia data layer is managed by the MILOS system
The Description Logic Component

For computational reasons, DL-MEDIA is based on a fuzzy 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

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

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

\[
\exists[1, 2]\text{MyMetadata} \sqsubseteq \exists[1, 2]\text{HasTitle}
\]
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 $\text{Person}(\text{firstname}, \text{lastname}, \text{age}, \text{email}, \text{sex})$

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

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

  $\exists[2, 1, 4]\text{Person}.((3 \geq 18) \land (5 = \text{'male'}) \sqsubseteq \exists[1, 2, 3]\text{AdultMalePerson}$
Examples (axioms involving $n$-ary relations)

- We also allow to specify textual and image similarity conditions

\[
(\exists[1] \text{ImageDescr}.([3] \sim \text{Img } \text{urn1}))) \land (\exists[1] \text{Tag}.([2] = 'sunrise')) \sqsubseteq \text{Sunrise}_\text{On}_\text{Sea}
\]

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

where $\text{urn1}$ identifies the image
A DL-MEDIA query consists of a conjunctive query of the form

$$q(x)[\text{score}] \leftarrow R_1(z_1), \ldots, R_l(z_l), \text{score} = f(\ldots),$$

$x$ is a vector of variables, and every $z_i$ is a vector of constants, or variables, $f$ score combination function.

- $q(x)[s] \leftarrow \text{Sunrise\_On\_Sea}(x)[\text{score}], s = \text{score}$
  // find objects about a sunrise on the sea

- $q(x)[s] \leftarrow \text{CreatorName}(x, y), (y = \text{‘paolo’}), \text{Title}(x, z), (z \sim \text{Txt ‘tour’})[\text{score}], s = \text{score}$
  // find images made by Paolo whose title is about ‘tour’

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

- $q(x)[s] \leftarrow \text{ImageObject}(x), \text{Int}(x, y_1)[\text{score}_1], \text{Car}(y_1), \text{Int}(x, y_2)[\text{score}_2], \text{Racing}(y_2), s = \text{score}_1 \cdot \text{score}_2$
  // find image objects about cars racing
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 $r(q, \mathcal{O})$
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
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 \( q_{\text{ref}}' \)
  - the number of reformulated queries after redundancy elimination \( q_{\text{ref}} \)
  - the time of the reformulation process \( t_{\text{ref}} \)
  - the number of queries effectively submitted to MILOS \( q_{\text{MILOS}} \)
  - the query answering time of MILOS for each submitted query \( t_{\text{MILOS}} \)
  - the time of merging process using the DTA \( t_{\text{DTA}} \)
  - the time needed to visualize the images in the user interface \( t_{\text{Img}} \)
  - the total time from the submission of the initial query to the visualization of the final result \( t_{\text{tot}} \)
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>
Conclusion

- Starting from a basic logic-based MIR model we have described some issues related to
  - Logic-based multimedia data categorization/annotation
  - Logic-based multimedia data retrieval
- While the Logic-based MIR is pretty old, we are still on its infancy
- However, now we have much more tools, standards and data
Elisa Bertino, Fausto Rabitti, and Costantino Thanos.
Multos: a document server for distributed office systems.

Nawei Chen and Dorothea Blostein.

Claudia d’Amato, Nicola Fanizzi, and Floriana Esposito.
A dissimilarity measure for ALC concept descriptions.

Horst Eidenberger.
Distance measures for mpeg-7-based retrieval.

Norbert Fuhr and Thomas Rölleke.
Hyspirit –a probabilistic inference engine for hypermedia retrieval in large databases.

Thomas Lukasiewicz and Umberto Straccia.
Tutorial: Managing uncertainty and vagueness in semantic web languages.
In *Twenty-Second Conference on Artificial Intelligence (AAAI-07)*, 2007.
Managing uncertainty and vagueness in description logics for the semantic web.

**Sherry Marcus and V. S. Subrahmanian.**
Foundations of multimedia database systems.

**Carlo Meghini, Fabrizio Sebastiani, and Umberto Straccia.**
A model of multimedia information retrieval.

**IEEE MultiMedia.**
MPEG-7: The generic multimedia content description standard, part 1.

**Fabrizio Sebastiani.**
Machine learning in automated text categorization.

**Umberto Straccia.**
PhD thesis, Department of Computer Science, University of Dortmund, Dortmund, Germany, June 1999.

**Umberto Straccia.**
Basic concepts and techniques for managing uncertainty and vagueness in semantic web languages.
Invited Lecture.

**Umberto Straccia.**
Managing uncertainty and vagueness in description logics, logic programs and description logic programs.

---

Umberto Straccia
Multimedia Retrieval and Reasoning
Umberto Straccia and Giulio Visco.
DLMedia: an ontology mediated multimedia information retrieval system.

Cornelis J. van Rijsbergen.
A non-classical logic for information retrieval.

Knowledge-assisted content-based retrieval for multimedia databases.