Multimedia Retrieval and Reasoning

Umberto Straccia

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

Outline

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
  - [vR86] C. J. van Rijsbergen. A non-classical logic for information retrieval. The Computer Journal, 29, pages 481-485, 1986.
    - Based on the estimation of the degree of implication between document d and query q, $d \rightarrow q$
  - [YKHI94]. A. Yoshitaka and S. Kishida and M. Hirakawa and T. Ichikawa. Knowledge-Assisted Content-Based Retrieval for Multimedia Databases. IEEE Multimedia, pages 12-20, 1994.
- Some good sources to previous work (most ideas have been addressed already):
  - [MS96]. S. Marcus and V. S. Subrahmanian. Foundations of Multimedia database Systems. Journal of the ACM, 43(3), pages 474-523, 1996.
  - [MSS01]. C. Meghini and F. Sebastiani and U. Straccia. A model of multimedia information retrieval. Journal of the ACM, 48(5), pages 909-970, 2001.
  - 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/`).

Basic MIR Model

- 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.

- 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

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

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 $\mathcal{R}$
  - 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, *i.e.*

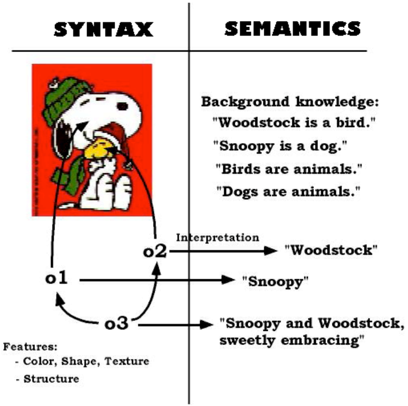
$$L_{Doc} = \langle L_{Syntax}, L_{Semantics}, Int \rangle$$

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

- Interpretation function
  - bridge between the syntax dimension and the semantics dimension determining the semantical meaning of the relevant objects identified at the syntax level

- Logic-based MIR
  - $L_{Semantics}$ and L_{Query} are logics
  - Integration between L_{Syntax} and $L_{Semantics}$ is defined in logical terms
  - $\mathcal{R}$ can be defined in terms of logical entailment

Example


SYNTAX
SEMANTICS
Background knowledge:
"Woodstock is a bird."
"Snoopy is a dog."
"Birds are animals."
"Dogs are animals."
Interpretation
o2
"Woodstock"
o1
"Snoopy"
o3
"Snoopy and Woodstock,
sweetly embracing"
Features:
- Color, Shape, Texture
- Structure

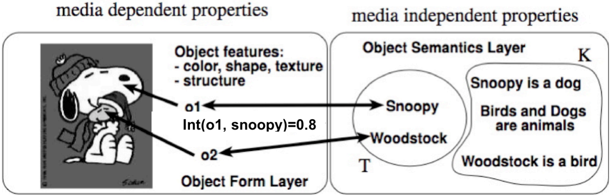
- Interpretation functions are, in general,
  - subjective
  - imprecise

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

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

- Automatic interpretation (annotation/classification) is currently the most difficult task

Basic MIR Model Ingredients



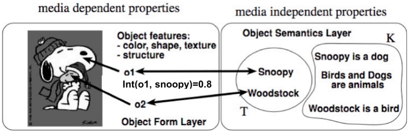
$L_{Doc} = \langle L_{Syntax}, L_{Semantics}, Int \rangle$

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

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

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

- $Int : Region \times Individual \rightarrow [0, 1]$
  - E.g., $Int(o_1, snoopy) = 0.8$ means: "region o_1 is about snoopy to degree 0.8"
- $L_{Semantics}$: a logic + Concrete domains supporting access to L_{syntax}



- L_{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”

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

$$
\begin{aligned}
Query(x, score) \quad \leftarrow \quad & ImageRegion(x), HasColor(x, white, score_1), \\
& Int(x, y, score_2), Animal(y), score = score_1 \cdot score_2
\end{aligned}
$$

- $HasColor(x, white, score_1)$: syntax-based query part ($score_1$ = degree of being region x white)
- $Int(x, y, score_2), Animal(y)$: semantics-based query part ($score_2$ = 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

$$Int\colon L_{Syntax} \times L_{Semantics} \to [0, 1]$$

- 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 $\Phi\colon L_{Syntax} \times L_{Semantics} \to [0,1]$, 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

$$
\begin{array}{lcl}
\Phi(o, bush) & \mapsto & 0.9 \\
\Phi(o, house) & \mapsto & 0.8 \\
\Phi(o, grass) & \mapsto & 0.7 \\
\Phi(o, tree) & \mapsto & 0.6
\end{array}
$$

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: $Int(o,c) \in \{0,1\}$

Image Classification Basics (survey, see [CB07])

- 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 a a-priori category assignment $\tilde{Int} \colon TS \times C \to \{0, 1\}$
  - $\tilde{Int}(o, c) = 1$ means that training image region $o \in TS$ has been (manually) assigned to category c
- Given $Top_k \colon Region \to TS$, where $Top_k(o)$ is the set of k regions $o_i \in TS$ that maximize $RSV(o, o_i)$
- Category Status Value (CSV) of $o \in Region$ w.r.t. category $c \in C$:

$$CSV(o, c) = \sum_{o' \in Top_k(o)} RSV(o, o') \cdot \tilde{Int}(o', c)$$

- 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

$$CSV(o,c) = \sum_{o' \in Top_k(o)} RSV(o,o') \cdot \tilde{Int}(o',c)$$

- $RSV(o,o')$ may be based on
  1. Visual low-level features in L_{Syntax} only (*e.g.*, MPEG7 standard distance measures [Eid03])
     $$RSV(o,o') = RSV_{Syntax}(o,o')$$
  2. Semantics properties of $L_{Semantics}$ only (use some semantic similarity function, *e.g.* [dFE06]),
     $$RSV(o,o') = RSV_{Semantics}(o,o')$$
  3. The combination of both, *e.g.*
     $$RSV(o,o') = \alpha \cdot RSV_{Syntax}(o,o') + (1-\alpha) \cdot RSV_{Semantics}(o,o'), \quad \alpha \in [0,1]$$
- In cases 2. and 3., logic-based reasoning comes into play, as usually semantic similarity functions rely on some logical reasoning
- We a have a seamless integration of feature-and logic-based classification

$$CSV(o,c) \quad = \quad \sum_{o' \in Top_k(o)} RSV(o,o') \cdot \tilde{Int}(o',c)$$

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)

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

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

$$\begin{aligned} CSV_{Semantics}(o, sportcar) &= \\ Pr(SportCar|AudiTT) &= Pr(AudiTT|SportCar) \cdot Pr(SportCar) \cdot (1/Pr(AudiTT)) \\ &\approx \frac{Pr(speed \geq 243|SportCar) \cdot Pr(accel \leq 6.9|SportCar) \cdot Pr(SportCar)}{Pr(speed \geq 243) \cdot Pr(accel \leq 6.9)} \end{aligned}$$

$$Pr(speed \geq 243|SportCar) =$$

$$\frac{|\{o \in TS \mid KB \models \exists y \exists z.Region(o) \wedge Int(o,y) \wedge Sportcar(y) \wedge hasSpeed(y,z) \wedge z \geq 243\}|}{|TS|}$$

$$\vdots$$

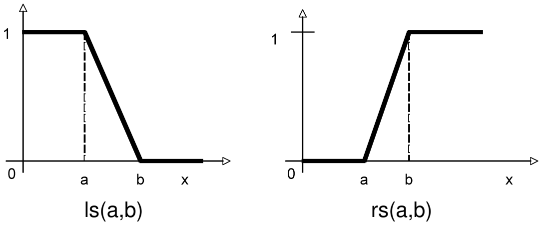
Example cont.

- Sport Car:

$$\forall x, hp, sp, ac\ SportCar(x) \iff 0.3HP(x, hp) + 0.2Speed(x, sp) + 0.5Accel(x, ac)$$

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

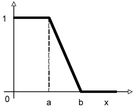
$$\begin{array}{rcl} HP(x, hp) & = & rs(180, 250)(hp) \\ Speed(x, sp) & = & rs(180, 240)(sp) \\ Accel(x, ac) & = & ls(6.0, 8.0)(ac) \end{array}$$



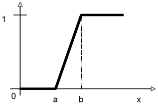
- Degree of truth of $SportCar(AudiTT)$: $0.1 \cdot 0.28 + 0.3 \cdot 1.0 + 0.6 \cdot 0.55 = 0.658$

- The fuzzy membership functions, as well as the weights, can be learned from a training set (large literature)

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



ls(a,b)



rs(a,b)

- Learned Training Sport Class:

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

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

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

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

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 $CSV_{Syntax}(o, c_i)$ and $CSV_{Syntax}(o_j, c_k)$
  - 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)

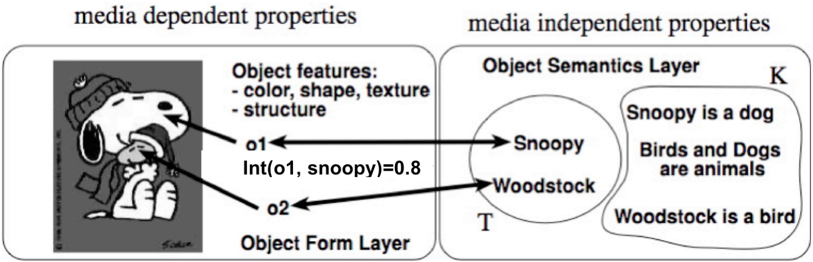
In my humble opinion, I encourage to rely on well grounded categorization methods (due to their success) rather than on method based on the first item. This latter idea is very old (see, e.g. [BRT88]) and suffers on effectiveness and efficiency (yet)

Logic-based Multimedia Retrieval

- 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 $\mathcal{R}$ function

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

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:

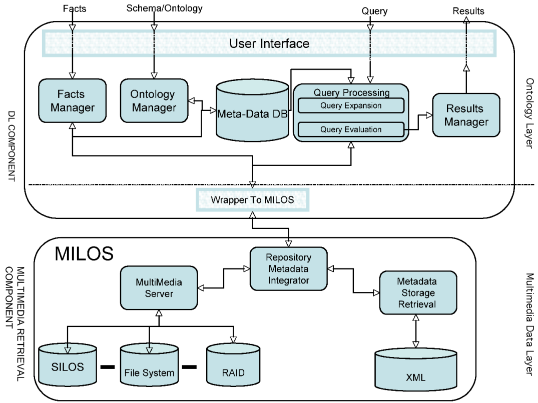
```
for $a in /library//pictures
where $a/ColourDistribution ≈ '...'
return $a/location
```

- 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”

$$\mathrm{ItalianCity} \sqsubseteq \mathrm{EuropeanCity}$$

- “Any italian city, which is also big is a big european city”

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

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)

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

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

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

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

DL-MEDIA Query Language

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

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

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

q(x)[s] $\leftarrow$ Sunrise_On_Sea(x)[score], $s = score$
// find objects about a sunrise on the sea

q(x)[s] $\leftarrow$ CreatorName(x, y), (y =' paolo'), Title(x, z), (z simTxt 'tour')[score], $s = score$
// find images made by Paolo whose title is about 'tour'

q(x)[s] $\leftarrow$ ImageDescr(x, y), (y simImg urn2)[score], $s = score$
// find images similar to a given image identified by *urn*2

q(x)[s] $\leftarrow$ ImageObject(x), Int(x, y_1)[$score_1$], Car(y_1), Int(x, y_2)[$score_2$], Racing(y_2), s = $score_1 \cdot score_2$
// find image objects about cars racing

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 $r(q, \mathcal{O})$
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

- 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.
In *Proceedings of the International Conference on Extending Database Technology (EDBT-88)*, number 303 in Lecture Notes in Computer Science, pages 606–615. Springer Verlag, 1988.

Nawei Chen and Dorothea Blostein.
A survey of document image classification: problem statement, classifier architecture and performance evaluation.
Int. J. Doc. Anal. Recognit., 10(1):1–16, 2007.

Claudia d'Amato, Nicola Fanizzi, and Floriana Esposito.
A dissimilarity measure for ALC concept descriptions.
In *SAC '06: Proceedings of the 2006 ACM symposium on Applied computing*, pages 1695–1699, New York, NY, USA, 2006. ACM.

Horst Eidenberger.
Distance measures for mpeg-7-based retrieval.
In *MIR '03: Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval*, pages 130–137, New York, NY, USA, 2003. ACM.

Norbert Fuhr and Thomas Rölleke.
Hyspirit –a probabilistic inference engine for hypermedia retrieval in large databases.
In *Proceedings of the 6th International Conference on Extending Database Technology (EDBT),*, Lecture Notes in Computer Science, pages 24–38. Springer-Verlag, 1997.

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

Thomas Lukasiewicz and Umberto Straccia.

Managing uncertainty and vagueness in description logics for the semantic web.
Journal of Web Semantics, 6:291–308, 2008.

Sherry Marcus and V. S. Subrahmanian.
Foundations of multimedia database systems.
Journal of the ACM, 3(43):474–523, 1996.

Carlo Meghini, Fabrizio Sebastiani, and Umberto Straccia.
A model of multimedia information retrieval.
Journal of the ACM, 48(5):909–970, 2001.

IEEE MultiMedia.
MPEG-7: The generic multimedia content description standard, part 1.
IEEE MultiMedia, 9(2):78–87, 2002.

Fabrizio Sebastiani.
Machine learning in automated text categorization.
ACM Computing Surveys, 34(1):1–47, 2002.

Umberto Straccia.
Foundations of a Logic based approach to Multimedia Document Retrieval.
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.
In *Reasoning Web, 4th International Summer School*, 2007.
Invited Lecture.

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

In *Reasoning Web, 4th International Summer School, Tutorial Lectures*, number 5224 in Lecture Notes in Computer Science, pages 54–103. Springer Verlag, 2008.

Umberto Straccia and Giulio Visco.
DLMedia: an ontology mediated multimedia information retrieval system.
In *Proceeedings of the International Workshop on Uncertainty Reasoning for the Semantic Web (URSW-08)*, 2008.

Cornelis J. van Rijsbergen.
A non-classical logic for information retrieval.
The Computer Journal, 29:481–485, 1986.

A. Yoshitaka, S. Kishida, M. Hirakawa, and T. Ichikawa.
Knowledge-assisted content-based retrieval for multimedia databases.
IEEE Multimedia, pages 12–20, 1994.