Multimedia Reasoning with Natural Language Support*

Stamatia Dasiopoulou
Multimedia Knowledge Group
Informatics and Telematics Institute
dasiop@iti.gr

Johannes Heinecke
France Télécom R&D
F-22307 Lannion cedex
johannes.heinecke@orange-ftgroup.com

Carsten Saathoff
Information Systems and Semantic Web
Universität Koblenz-Landau
http://isweb.uni-koblenz.de/
saathoff@uni-koblenz.de

Michael G. Strintzis
Information Processing Laboratory
Aristotle University of Thessaloniki
strintzi@eng.auth.gr

Abstract

In this paper we present an approach that combines multimedia reasoning and natural language processing for the semantic integration of automatic and manual image annotations based on domain ontologies. We discuss how to apply natural language processing to transform natural language descriptions and queries into an ontological representation that allows users to formulate formal semantics in an intuitive manner, without the need to cope with complex ontological structures and unwieldy user interfaces. Illustrative experimental examples demonstrate the added value.

1. Introduction

The amount of digital media is growing rapidly and multimedia content is omnipresent. However, little progress has so far been made in indexing, organizing or retrieving the available content in an efficient way that reflects user needs. While activities such as TrecVid[1] show that well performing analysis algorithms can be built, these are usually focused on solving narrow problems that address constrained datasets. As a result, their application in real life scenarios is rather questionable due to the immense effort required to generate a sufficiently large number of concept detectors that work well on practically arbitrary content, and the serious challenges in avoiding the performance degradation witnessed with the increase of supported concepts. On the other hand, progress has been made in analysis approaches for the detection of rather generic concepts such as faces and persons, as well as coarse content classification such as natural vs. manmade. Combined in a synergistic approach with manual annotation, such efforts have the potential to achieve improved indexing and retrieval.

In this paper we describe an approach that combines *Multimedia Reasoning* and *Natural Language Processing (NLP)* for enhancing automatically produced annotations and providing natural language queries for semantic enabled retrieval. More specifically, the core contribution of the proposed approach constitutes the mapping of manual natural language annotations (descriptions) to an ontological representation, and their subsequent integration with automatically produced annotations based on a set of domain ontologies through the utilization of reasoning methodologies. The ontological representation of natural language entails precise semantics, while reasoning ensures the semantic coherency of the annotations and their further enrichment. Enabling the use of the same domain conceptualization, the manual descriptions serve as background information for correctly interpreting the automatic analysis results, while the user can enjoy the benefits of ontology-based retrieval through intuitive means.

The currently deployed system architecture is shown in Fig. 1. An input image is first analysed by a set of image analysis modules. The only requirement is that the annotations resulting from the analysis modules adhere to the domain ontologies; no assumptions need to be made about the concrete implementation. The automatically produced annotations are subsequently augmented with user entered textual descriptions that are transformed into an ontological representation adhering to the employed domain ontologies. As the knowledge representation formalism we chose RDF to benefit from the explicit semantics and existing tools sup-

*This work was supported by the European Commission under contract FP6-001765 aceMedia (*http://www.acemedia.org/*).

[1] *http://trecvid.nist.gov/*

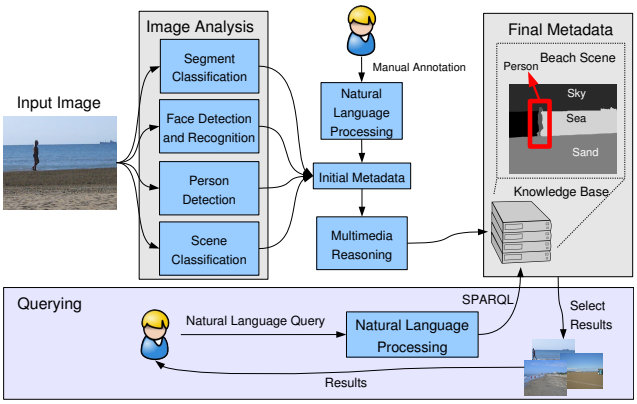

Figure 1. The aceMedia architecture

port, while OWL-DL statements has been included where additional expressivity was required. In the sequel, the initial set of annotations undergoes a two-stage reasoning process in order to obtain a semantic coherent, possibly enhanced final annotation, which is eventually stored in a RDF repository that allows querying the content using SPARQL. For the retrieval, the user posed natural language queries are transformed into corresponding SPARQL queries, utilising the same approach developed for handling the manual textual descriptions. The transformation of natural language to an ontological representation for both description and querying provides several benefits, namely: (i) a natural way of formulating descriptions and queries, (ii) automatic mapping of synonyms to the correct ontological classes, (iii) correction of spelling errors exploiting the semantics of remaining input, and (iv) straightforward expression of what would constitute complex ontology-based queries, through the automatic mapping of relations expressed in natural language to their ontological counterparts.

The remainder of the paper is structured as follows. Section 2 shortly demonstrates how the proposed approach can be deployed in a real application. The developed natural language processing methodology is described in Section 3, addressing both the ontological mapping and the transformation of linguistic data, while Section 4 describes the developed reasoning modules. In Section 5, exemplar queries demonstrate the benefits brought. The paper concludes with a short discussion on related work and future research directions in Sections 6 and 7, respectively.

2. Application Scenario

Mary loves to take images of her family during different occasions, such as holidays or birthdays. For managing her images she uses the aceMedia system, which offers automatic analysis of content, means to manually enter natural language descriptions, and natural language queries for retrieval. When Mary uploads images, the system starts the automatic analysis and produces annotations. In parallel, or at a later time, Mary can manually annotate her images. She can either annotate images with higher-level level information, usually not found automatically, such as location and special events names, or she can add some text to images that she particularly likes. Utilising the proposed combined natural language processing and multimedia reasoning framework, the system integrates her manual annotations with the automatically produced ones into a final consistent annotation, and offers the means to query the integrated metadata in a coherent manner.

For instance, her two children John and Jane both like to play softball. Recently she was searching for some image of her children playing softball on the beach during last years vacation in Greece. She enters a query “John and Jane playing softball on the beach”, and the system returns a list of images with John and Jane playing softball on the beach. However, she never had to annotate these images in the same way as she queried for them. In fact, she only annotated these images with “John and Jane playing softball”, as this kind of information can not be detected automatically. The information that the images were taken on the beach was then automatically added by the system and integrated with her manual input. For this reason, other images that depict her children playing softball in a school tournament or in their backyard are not returned.

3. Natural Language Processing for Ontological Reasoning

In order to be able to create ontological representations of textual data in ontology languages such as RDF and OWL, or query languages like SPARQL, the semantics of natural language must be mapped onto the classes and relations (properties) of the corresponding ontology. In an ideal world this should be done automatically in order to provide effective adaptability to new domains (i.e. new ontologies). The research undertaken in aceMedia has resulted in a nearly automatic mapping, described in the following.

3.1. Mapping of Linguistic Data and Ontologies

As will be detailed below in section 3.2, a deep analysis approach has been followed in order to transform linguistic expressions into ontological representations. This means we do not only have a complete lexicon and (dependency) grammar rules but also a semantic thesaurus which defines the meanings of lexical entries. The task of the mapping is to establish correspondences between the ontological classes and relations and the entries of the semantic thesaurus; cardinality and property restrictions as

well as class constructors are currently ignored. The semantic thesaurus is completely language independent even though it may contain meaning definitions, which are not lexicalised in some languages or just in a single one. Currently our English and French lexica (several ten thousand lexicon entries for each language [5]) are linked extensively with the semantic thesaurus which contains about a 100 000 definitions[2]. This means that the result of the mapping is also language independent. The semantic thesaurus provides the following information: (i) a *semantic predicate*[3] including arguments and sortal restrictions for these arguments for every lexicon entry (i.e., this means that homonyms with different meanings have different predicates as well as synonymic entries share identical predicates), and (ii) a *thematic hierarchization* of these predicates, i.e. a grouping of different predicates into themes, domains and super-domains. These predicates are used to build semantic graphs, which express the meaning of an analysed sentence.

Even though our semantic thesaurus contains some information on synonyms, it is not possible to define "being synonym" in a general and domain independent way. For instance, in the aceMedia domains there is only a single class for (four-wheel) vehicles. In consequence, lemmas like *car*, *lorry*, *vehicle*, *truck* etc. need to be considered as synonyms, since their differences are ignored by the ontology. However, in a different application domain, e.g., the car manufacturing domain, these may as well be distinctive. such domain-specific synonyms are added to the semantic thesaurus prior to the mapping[4] to be exploitable.

In order to be mappable, the classes of the ontologies should correspond to expressions in natural language, e.g. *player*, *mountain shelter*, *line judge*. However, classes which model complex composite notions, such as *Post_Office_With_Late_Night_Opening_Hours*, are not yet sufficiently mapped, while (ontological) relations which correspond to (transitive) verbs in natural language such as *isWonBy* and *contains* are the easiest to map. Of course we are aware of the fact that ontologies not only contain relations which correspond semantically to transitive verbs. Especially in relations with a literal range, the mapping poses additional particularities. Whereas string ranges, e.g., for relations such as *hasName* and *hasNationality*, are currently mapped by the automatic process, integer and boolean ranges can be trickier. Although relations with a boolean range only can have two different values, in natural language these values can be hidden in various constructions. For example *tennis:isCovered* is the relation to express whether a stadium is open-air or indoors. This means that linguistic expressions like *to cover*, *indoors*, *open air*, *has a roof*, and their negated counterparts, need to be mapped onto the considered relation.

```
:V286681
    a      midlevel:boat .
:V286672
    a      dolce:Natural-Person ;
    midlevel:hasFirstName  "Nicolas" ;
    midlevel:isRightOf   :V286681 .
```

Example 1. Ontological representation (N3)

Apart from the ontologies, the mapping requires a complete lexicon of the language used to label or describe the ontological classes and relations, which is linked to a semantic thesaurus. The ontologies, on the other hand, need non-ambiguous class and relation names (like *tennis:Player*)

The ontologies used in aceMedia are based on the DOLCE [10][5] ontology. This is important since the taxonomical structure of aceMedia's domain ontologies therefore depends on the DOLCE ontology.

The mapping itself comprises several steps, summarised in the following (for a more detailed description the reader is referred to [6]): (i) extraction of the "ontological context" of classes and relations and assignment of eventual reformulations of class/relation names, (ii) detection of classes meaning using their ontological context, i.e., direct subclasses, (iii) detection of relations meaning using the corresponding domain and range ontology definitions, (iv) addition of the ontological hierarchy to the semantic taxonomy, (v) creation of semantic transformation rules for so called *complex class names*[6], and (vi) creation of transformation rules for the acquisition of ontological representation from semantic graphs (cf. section 3.2 on semantic graphs).

3.2. Transforming Natural Language into Ontological Representations

The objective is to make the semantics of textual descriptions explicit, thus available for formal processing like reasoning and semantic querying. For instance, a description like *Nicolas is at the right of the boat* should yield the RDF statements given in Example 1.

For textual descriptions, as well as for user queries, several methods of processing can be considered, ranging from

[2] Other languages such as German, Arabic, Spanish and Chinese are partially linked. Further work is in progress.

[3] A predicate is defined as the semantic concept and its arguments. Note that words in different languages but with the same meaning are linked to the same predicate.

[4] Currently this is a semi-automatic process, since the definition of synonyms is highly domain and application specific.

[5] Cf. also *http://www/loa-cnr.it/Papers/dolce_docs.zip*.

[6] By this term we refer to names of ontology classes (and, of course, relations) which use multi-word expressions or phrases like *tennis:ExhibitionMatch*.

very simple key word spotting algorithms to more or less profound linguistic analysis. In the presented approach, we opted for a deep syntactic and semantic analysis in order to be able to detect semantic relations between actants, i.e. to be able not only to transform textual descriptions into a list of ontological classes and instances, but also to exploit syntactic and semantic information and establish the ontological relations linking these instances. In other words, ontologies provide ontological classes, instances of which have to be identified within the natural language input. But the ontologies provide also relations between classes. These relations are also expressed in natural languages: as it happens, this is frequently done by verbs, adjectives and prepositions. A simple keyword- (or class-) spotting mechanism can detect the relations expressed in a text, but it is unlikely it will be able to find the relata, i.e. the domain and the range of the ontological relation. In order to find these, a deep analysis that can build a semantic representation is required (cf. [7]). This semantic representation is then transformed into an ontological representation. Exploiting the associations of the previous step (linguistic-ontological mapping, section 3.1) that link a semantic concept (or graph of semantic concepts) to every ontological class and relation, the final ontological representation is acquired.

Another important argument justifying for a semantic analysis is the benefits arising from synonyms and antonyms exploitation (section 3.1). This is further illustrated when one is comparing the semantic richness of natural language with the usually "restricted" model of the world as presented by (domain) ontologies. For instance, the domain ontologies used in aceMedia cover the domains holidays (sea, mountains, camping), family as well as tennis and motorsports. Whereas our ontologies have one concept for *beach*, *cliffs*, *mountain* and *sea*, natural languages (in our case English and French) provide several expressions for each of these ontological classes.

A final argument for employing such an approach is the possibility of detecting and correcting errors in the textual input (descriptions or user queries), supporting typographical and diacritical (for languages like French and German) errors in otherwise inanalysable input. Furthermore, as a last resort a correction based on phonetic similarity is available.

Currently, the proposed approach is based on a syntactic analysis using a dependency grammar (cf. [17, 11]), resulting in the translation of phrases into dependency syntax trees. There are several other advantages of a deep analysis approach. In our case, the syntax analyser is very robust in order to tackle ungrammatical input. Furthermore, there is a context dependent typographic correction. In the worst case, partial syntactic trees are obtained for every chunk that could be identified within the textual input, i.e., partial syntactic trees, containing just one word, may result. Furthermore, in case of ambiguity at the lexical level, a context driven choice of the semantic concept is employed. Thus, starting from the syntax tree we are able to create semantic graphs which are composed of semantic predicates. We are now able to transform the semantic representation into an ontological representation (as shown above, cf. Ex. 1).

```
:V5  a      tennis:Player ;
     tennis:hasFirstName  "André" ;
     tennis:hasName  "Agassi" .
:V4  a      tennis:Trophy ;
     tennis:isWonBy  :V5 .
```

Example 2. Incoherent ontol. representation

```
:V12750
     a      tennis:Player ;
     tennis:hasFirstName  "André" ;
     tennis:hasName  "Agassi" .
:V12749
     a      tennis:Trophy .
:V12751
     a      tennis:Finale ;
     tennis:isWonBy  :V12750 .
:N1  a      tennis:Tournament ;
     tennis:awards  :V12749 ;
     tennis:hasFinale :V12751 .
```

Example 3. Relinked instance of *tennis:Trophy*

Another important aspect is the implementation of an "ontological correction" or "adaptation", which ensures that the ontological representations created from natural language are coherent with aceMedia's (domain) ontologies. For example, a sentence like *the trophy is won by Agassi* results initially in the ontological representation shown in Ex. 2. This very ontological representation, however, has a flaw: it is not necessarily coherent with the underlying ontology, which in our case states that: (i) only *Finals* can be won (by *Players*), i.e. the domain of the relation *tennis:isWonBy* is the class *tennis:Final*, which in turn is not a super-class of *tennis:Trophy*; (ii) *Trophies* are *awarded* by *Tournaments*; and (iii) *Tournaments* have *Finals*.

In order to obtain an ontologically coherent representation, the incorrectly linked class (*tennis:Trophy*) is separated from the relation *tennis:isWonBy* and replaced by the class defined as domain for the relation *tennis:isWonBy* in the ontology.

In a final step, we try to relink the now isolated instance with the rest of the ontological representation by inserting further instances and relations, if the ontologies allows us to do so. The way to achieve this is defined in the ontology. We thus arrive at the corrected representation (Ex. 3). After this ontological correction the generated representation is coherent with the ontology. It has to be noted that the

```
:D45317
    a       midlevel:unknown ;
    midlevel:isOfType   ”object.drink.juice” .
:D45318
    a       midlevel:unknown ;
    midlevel:isOfType   ”object.fruit.orange” .
```

Example 4. Ontol. representation of semantic predicates not represented by ontol. classes

relinking is not done for the SPARQL queries in order not to complicate the queries and possibly block the retrieval.

Finally we process the frequent case where textual content providers or users issuing queries, unaware of the underlying ontologies, produce utterances which employ semantic concepts which are not within the scope of the domain ontologies; e.g. the aceMedia application has (amongst others things) a tennis ontology. A query on *orange juice*, however, cannot be translated into a valid ontological representation, since there is no ontological class which corresponds to the predicate of the thesaurus for the notion “orange juice”. Instead of creating an empty representation, a special ontological class may be provided by the ontology, namely a *midlevel:unknown* class which has a relation *midlevel:isOfType*. This relation has a string literal as range, which will hold the name of predicate which is without corresponding class. Therefore a phrase like *orange juice* results in an ontological representation such as Ex. 4. A SPARQL-query uses the same values for the *midlevel:isOfType*-relation and will therefore be able to retrieve the needed information from a knowledge base even if the concepts of both, the description and the user query, are not covered by the ontology. The final SPARQL query is sent to the knowledge-base in order to retrieve the contents which metadata matches with the query.

For the retrieval, the choice of the same syntactic-semantic analysis of user queries entails the following benefits: (i) queries can enjoy the typographical corrections of the NLP module, including correction strategies which try to map unknown words to known words by typographical and phonetic correction, (ii) the processing of synonyms, due to the intermediary semantic representations, allows users to use the wide range of possible linguistic expressions, and (iii) since the ontological representations are language independent, multilingual queries are supported (currently English and French), independently of the language used for textual descriptions.

4. Inferring Consistent Semantic Metadata

As illustrated in Fig. 1, the automatically produced and manually entered annotations are combined into an initial semantic annotation that may introduce complementary, overlapping, or even contradictory descriptions. To reach the final content annotation, reasoning is applied to enforce the smooth integration of the initial annotations into a semantically coherent set, further enriched, if possible, with higher level descriptions. Due to the two intertwined levels in which annotations come, namely the segment and the scene level, the reasoning goals translate into the following tasks: (i) semantic consistency at segment level, (ii) semantic consistency checking at scene level, (iii) semantic coherency checking among scene and segment level annotations, and (iv) higher level descriptions inference.

Given the very nature of multimedia analysis, two principal requirements emerge, namely support for *uncertainty* and *provenance*. Uncertainty refers to the ambiguity inherent in multimedia analysis that causes the produced annotations to be characterised by a relative degree of confidence, reflecting their plausibility, while provenance accounts for the different media types and different principles that the available analysis modules work on and that affect their performance and reliability. Taking into consideration the aforementioned, we developed a reasoning methodology that combines fuzzy constraint (section 4.1) and fuzzy DL (section 4.2) reasoning to effectively meet the challenges involved.

4.1. Topological Reasoning with Constraints

The goal of the topological reasoning within our approach is to provide a consistent labelling of different regions within the image. In [13], the methodology to transform the image labelling problem into a *Constraint Satisfaction Problem (CSP)* is presented. To better account for the heuristics needed in image interpretation, this approach has been extended to *Fuzzy Constraint Satsifaction Problems (FCSP)*.

A FCSP [1] is a system of variables and constraints. Each constraint is defined as a relation on a number of variables, and effectively constrains the set of legal assignments to these variables. Each tuple of a constraint is accompanied by a degree that represents to what extend the given tuple satisfies the constraint, allowing thus to model preference among certain solutions. The fuzzy constraint reasoner then searches for the solution satisfying the posed constraints optimally.

To find an optimal labelling of the regions of an image, first the segment classification results are transformed into a FCSP. Each segment is represented by a variable. Spatial relations are extracted between any two segments, and are added as constraints to the resulting FCSP. In the background knowledge, for each spatial relation, the valid combinations of concepts are defined, i.e., the relation that defines a constraint is modelled explicitly. In case of the *above* relation for example, sky is allowed to be depicted above the

$$\text{Natural} \equiv \text{Outdoors} \sqcup \neg\ \text{ManMade}$$
$$\text{Cityscape} \sqsubseteq \text{Manmade}$$
$$\text{Beach} \equiv \exists\text{contains.Sea} \sqcap \exists\text{contains.Sand}$$
$$\text{Beach} \sqsubseteq \text{Natural}$$
$$(\mathit{image1} : \mathit{Cityscape})$$
$$(\mathit{region1} : \mathit{Sea})$$
$$(\mathit{region2} : \mathit{Sand})$$
$$(\mathit{image1}, \mathit{reqion1}) : \mathit{contains}$$
$$(\mathit{image1}, \mathit{reqion2}) : \mathit{contains}$$

Example 5. Example Outdoor TBox and ABox

sea, but not vice versa. A unary constraint on each segment variable is used to represent the degrees of confidence of each label produced by the segment classification. Finally, using standard branch and bound search, the final labelling is computed.

4.2. DLs Reasoning for Multimedia Annotation

Description Logics (DLs) [2] are a family of knowledge representation formalisms characterised by logically founded formal semantics and well-defined inference services. Starting from the basic notions of atomic concepts and atomic roles, where concept and role are the counterpart for class and properties in the ontology terminology, arbitrary complex concepts can be described through the application of corresponding constructors (e.g., $\neg$, $\sqcap$, $\forall$). Terminological axioms (TBox) allow to capture equivalence and subclass semantics between concepts and relations, while real world entities are modelled through concept ($a : C$) and role ($R(a,b)$) assertions (ABox). *Satisfiability*, *subsumption*, *equivalence* and *disjointness* constitute the TBox inference services, and *consistency* and *entailment* the ABox ones.

Consequently, given a TBox that describes a specific domain, using DLs one can build an ABox from the available initial annotations and benefit from the inferences provided to detect inconsistencies and obtain more complete descriptions. Assuming for instance the TBox and ABox of Ex. 5, the following assertions entail: (*image1* : *Cityscape*), (*image1* : *ManMade*), (*image1* : *Mountain*), and (*image1* : *Natural*). Furthermore an inconsistency is detected, caused by disjointness axiom relating the concepts Natural and ManMade. However the aforementioned apply only to the case of crisp assertions (annotations). In order to handle the uncertainty of analysis, the extension of the DLs semantics is needed. Two main efforts exist currently that address formally both the semantics and the corresponding reasoning algorithms for fuzzy DL extensions. In [15], the so called f-*SHIN* is introduced which extends the semantics of concept and role assertions. More specifically, an f-*SHIN* ABox consists of a finite set of *fuzzy assertions* of the form $a : C \bowtie n$ and $(a,b) : R \bowtie n$, where $\bowtie$ stands for $\geq$, $>$, $\leq$, and $<$[7]. In [16], a fuzzy extension of *SHIF(D)* is presented, continuing earlier works on *ALC* and *SHIF*. In addition to the extended fuzzy semantics, the authors present a set of interesting features: (i) concrete domains as fuzzy sets, (ii) fuzzy modifiers such as *very* and *slightly*, and (iii) fuzziness in terminological axioms as well.

Extending DLs with fuzzy semantics allows to apply the inference services on fuzzy annotations as well. Let us augment the of the previous example with the axiom $\mathit{Swimmer} \equiv \mathit{Face} \sqcap \mathit{inside}.\mathit{Sea}$, and assume the following analysis produced annotations: $(\mathit{region1} : \mathit{Sea}) \geq 0.8$, $(\mathit{region2} : \mathit{Sand}) \geq 0.6$, $(\mathit{region3} : \mathit{Person}) \geq 0.7$, and $((\mathit{region3}, \mathit{reqion2}) : \mathit{inside}) \geq 0.6$. Utilizing the extended fuzzy DLs semantics, greatest lower bound values of 0.9 and 0.6 result for *image1* with respect to the concepts Natural and Beach respectively, and a value of 0.6 is obtained for *region3* with respect to the concept Swimmer. However, both initiatives, *FiRE*[8] and fuzzyDL[9], respectively, are in a quite early stage of implementation, providing query services limited to subsumption and greatest lower bound, that render them unsuitable for a realistic application.

In order to overcome the practical restrictions resulting from the current stage in fuzzy DLs implementations, we propose an extended DL-based reasoning approach that enables the use of existing crisp DL reasoners, while at the same time providing support for handling fuzzy assertions. Towards this end, the fuzzy membership values semantics have been defined in accordance with the fuzzy DLs extensions initiatives, while their actual handling, i.e. the readjustment of the corresponding values, has been implemented as a separate mechanism that interacts with the standard DL reasoner. Following such an approach is practically feasible due to the simplifications that hold within the examined multimedia analysis context. More specifically, since the considered roles are crisp, the fuzzy set semantics of $\exists R.C$ and $\forall R.C$ formulas depend solely on $C^I(a)$, which practically means that all considered membership values refer solely to fuzzy concept conjunction, disjunction and negation. For the corresponding fuzzy operations we follow the Łukasiewicz negation, $c_L(a) = 1 - a$, and the Gödel norms, $t_G(a,b) = min(a,b)$ and $u_G(a,b) = max(a,b)$.

Having defined the uncertainty handling methodology, the semantic coherency related reasoning tasks have been implemented as follows.

Semantic consistency checking at scene level. All analysis modules may be involved in this task, contributing scene

[7] Intuitively a fuzzy assertion of the form $a : C \geq n$ means that the membership degree of the individual a to the concept C is at least equal to n

[8] *http://www.image.ece.ntua.gr/~nsimou*

[9] *http://faure.isti.cnr.it/~straccia/software/fuzzyDL/fuzzyDL.html*

level annotations either explicitly or implicitly. To be able to include scene level assertions from segment level annotations, we ignore all disjointness axioms. Let K be the number of domain level concepts, N the number of analysis components involved, d_{ij} the degree of confidence produced by the component i for the domain j and w_i the weight reflecting the reliability of the module i. The domain concepts degree of confidence is updated successively, moving from more specific concepts to more generic ones, using the following formula: $CD_k = \sum_i w_i * d_{ik} + \sum_j w_j * d_{jl}$, where i, j range over the analysis components that have produced a degree of confidence for the respective concept, and l ranges over the concepts that are subsumed by the under consideration concept. The two sums are appropriately normalised so that the final degree is in $[0,1]$. At each hierarchy level, the concept with the higher degree amongst its (disjoint) siblings is kept.

Semantic coherency checking among scene and segment level annotations. Having inferred and appropriately adjust the confidence of the scene level annotations applicable to the examined content, the next step is to ensure the semantic coherency of object level assertions. To accomplish this task, disjointness axioms are exploited. Within the current set of supported concepts, the number of such axioms is quite restricted, as most of the supported concepts can be present in more that one of the considered sub-domains within the scene level concepts. The notion of *non-concept* needs to be introduced to allow the identification of annotations that caused the inconsistency and thus their removal from the corresponding annotation metadata.

Higher level descriptions inference. The inference of higher level annotations is the most straightforward of the considered tasks. As long as the means to handle uncertainty are provided, then the only prerequisite consists in the proper engineering of the domain knowledge. Apart from the DL terminological axioms, a set of DL safe rules has also been employed to allow inferring role assertions. The latter aims to cover annotations describing activities in a intuitive way, e.g. "person standing on the beach", being closer to human cognition is more likely to be entered as user query than the semantically equivalent, but verbose, "person above/left/right sand".

5. Results

In the following, we exemplify some of the advantages that the proposed approach entails through common query examples, and discuss how these would otherwise fail, i.e., in case the natural language ontological representation or the multimedia reasoning were missing. A very common category of queries are the ones that indicate some spatial relation among the concepts (objects, persons, etc.) of interest. Going back to the scenario of Section 2, Mary, wishing to retrieve all beach images from last year's vacations, where a cliff is depicted on the left, may enter a query like "cliffs at the left of the beach". In Fig. 2, the final metadata acquisition of such an example image is shown. Reasoning enforces the semantic consistency of the annotations, maps image segment spatial relations to domain concept relations, and infers "beach". The NLP translates the natural language query into an ontological representation and enables the retrieval of the image. Mary's cousin, Jean posing the (French) query "falaises à gauche de la plage" ("cliffs left to the beach"), can retrieve exactly the same images too, due to the language independent functioning of the NLP module. Another example of the advantages entailed by translating natural language into ontological representation refers to the flexible query formulation support. For instance an image of the tennis-domain with the textual description as "André Agassi wins the finals at Wimbledon" can be retrieved by "Agassi winning games".

The aforementioned constitute few indicative examples of the entailed added value. The application of reasoning improves on the initial automatic annotations, eliminating inconsistencies and further enriching them. Furthermore, initial evaluation showed clear benefits concerning both domain specific synonyms and multilingual support, while the robust syntactic analysis provides a certain "protection" against user errors, allowing to get a result even if the input (descriptions or user queries) are erroneous. The current limits are less of theoretical nature, but due to yet incomplete work on the mapping; as said above some relations, notably those with literal values as ranges, are currently ignored. Another minor issue, that however hinders the wide scale evaluation of the proposed system, concerns queries (descriptions) that are utterly out of the scope of the domain ontologies. As synonyms are only definable within a specific domain, the *midlevel:unknown*-class (and its *midlevel:isOfType* relation) use the semantic predicate, but with synonyms information missing.

6. Related Work

Mapping linguistic data onto ontologies is very closely related to to ontology alignment [8]. Due to space limitations, we only point to two interesting approaches to a similar problem. In order to create SPARQL queries from natural language, a simple incremental grammar which is ontology-driven is proposed in [3]. Thus, while the user types his query, the system proposes the vocabulary and the relations he can use. Obvious limitations of this approach for our needs include that user descriptions are not necessarily analysed at the time they are entered, while in addition, pre-existing textual descriptions, which have been formulated independently of the aceMedia-system and its ontologies need to be handled. Furthermore, such approach is quite restricted in terms of allowing descriptions and queries, which

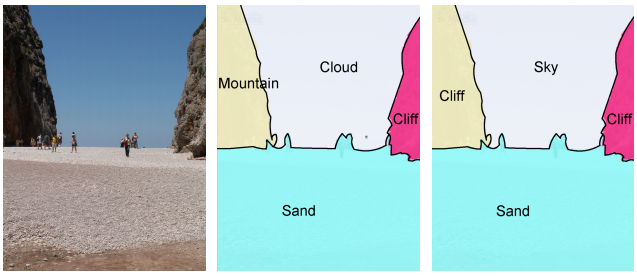

Figure 2. metadata acquistion

go beyond the "world" as modeled by the included ontologies. In [9], NLP is used to create triple-based linguistic queries, based on who/which/when-questions, which are then mapped onto ontological (RDF) triples by a Relation Similarity Service (RSS), using ontologies and linguistic resources like WordNet [4]. DLs and by extension ontologies have been employed in quite several approaches to semantic image analysis for the representation of the required knowledge. However, in the majority of the literature, focus remained on their exploitation as shareable vocabularies for representation, rather than as means for formal inference. Works adhering to the latter perspective include [14, 12], that consider segment level semantic descriptions.

7. Conclusion and Perspectives

In this paper we presented the aceMedia approach to semantic multimedia analysis that combines reasoning and natural language support in order to semantically integrate and enrich analysis results, while providing semantic enabled retrieval in natural language queries. Concerning the work on linking NLP with ontological reasoning, future work will concentrate on improving the linguistic-ontological mapping, and particularly regarding relations with XML datatype ranges. Further we envisage to automate to a certain degree the detection of synonyms for the mapping of ontological classes. Future plans with respect to the presented fuzzy reasoning framework include further investigation of the fuzzy expressivity required within the multimedia analysis context and of its formalization, including the impact of different fuzzy operators on the entailed semantics.

References

[1] K. R. Apt. *Principles of Constraint Programming*. Cambridge University Press, 2003.

[2] F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi, and P. F. Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, 2003.

[3] A. Bernstein and E. Kaufmann. GINO - A Guided Input Natural Language Ontology Editor. In *ISWC 2006*, Lecture Notes in Computer Science 4273, pages 144–157, Heidelberg, 2006. Springer.

[4] C. Fellbaum. *WordNet. An Electronic Lexical Database*. MIT Press, Cambridge, MA., 1998.

[5] E. Guimier de Neef, M. Boualem, C. Chardenon, P. Filoche, and J. Vinesse. Natural language processing software tools and linguistic data developed by France Télécom R&D. In *Indo European Conference on Multilingual Technologies*, Pune, India, 2002.

[6] J. Heinecke. Génération automatique des représentation ontologiques. In P. Mertens, C. Fairon, A. Dister, and P. Watrin, editors, *Verbum ex Machina. TALN 2006, vol. 2*, pages 502–511. Presses universitaires de Louvain, Louvain, 2006.

[7] J. Heinecke and F. Toumani. A Natural Language Mediation System for E-Commerce applications. In *Workshop Human Language Technology for the Semantic Web and Web Services, ISWC, Sanibel, Florida*, pages 39–50, 2003.

[8] Y. Kalfoglou and M. Schorlemmer. Ontology Mapping: The State of the Art. In Y. Kalfoglou, M. Schorlemmer, A. Sheth, S. Staab, and M. Uschold, editors, *Semantic Interoperability and Integration*, Dagstuhl Seminar Proceedings 4391, Schloß Dagstuhl, 2005.

[9] V. Lopez, M. Pasin, and E. Motta. AquaLog: An Ontology-Portable Question Answering System for the Semantic Web. In *ESWC 2005*, Lecture Notes in Computer Science 3532, pages 546–562, Heidelberg, 2005. Springer.

[10] C. Masolo, S. Borgo, A. Gangemi, N. Guarino, and A. Oltramari. Wonder Web Deliverable D18: Ontology Library. Technical report, 2003.

[11] I. A. Mel'čuk. *Dependency syntax. Theory and Practice*. State University Press of New York, Albany, 1988.

[12] B. Neumann and R. Möller. On scene interpretation with description logics. Technical Report FBI-B-257/04, 2004.

[13] C. Saathoff. Constraint reasoning for region-based image labelling. In *IEE International Conference of Visual Information Engineering (VIE)*, 2006.

[14] J. P. Schober, T. Hermes, and O. Herzog. Content-based image retrieval by ontology-based object recognition. In V. Haarslev, C. Lutz, and R. Möller, editors, *KI-2004 Workshop on Applications of Description Logics (ADL-2004)*, Ulm, Germany, September 2004.

[15] G. Stoilos, G. Stamou, V. Tzouvaras, J. Pan, and I. Horrocks. The fuzzy description logic f-SHIN. In *International Workshop on Uncertainty Reasoning For the Semantic Web*, Galway, Ireland, November 2005.

[16] U. Straccia. A fuzzy description logic for the semantic web. In E. Sanchez, editor, *Fuzzy Logic and the Semantic Web*, Capturing Intelligence, pages 73–90. Elsevier, 2006.

[17] L. Tesnière. *Éléments de syntaxe structurale*. Klincksieck, Paris, 1959.