

Annotating text using the Linguistic Description Scheme of MPEG-7: The DIRECT-INFO Scenario

Thierry Declerck, Stephan Busemann
Language Technology Lab
DFKI GmbH
Saarbrücken, Germany
{declerck|busemann}@dfki.de

Herwig Rehatschek, Gert Kienast
Institute for Information Systems &
Information Management,
JRS GmbH
Graz, Austria
{rehatschek|kienast}@joanneum.at

Abstract

We describe the way we adapted a text analysis tool for annotating with the Linguistic Description Scheme of MPEG-7 text related to and extracted from multimedia content. Practically applied in the DIRECT-INFO EC R&D project we show how such linguistic annotation contributes to semantic annotation of multimodal analysis systems, demonstrating also the use of the XML schema of MPEG-7 for supporting cross-media semantic content annotation.

1 Introduction

In the R&D project DIRECT-INFO the concrete business case of sponsorship tracking was targeted. The scenario investigated within the project was that sponsors want to know how often their brands are mentioned in connection with the sponsored company. The visual detection of a brand (e.g. in videos) is not sufficient to meet the requirements of this business case. Multimodal analysis and fusion – as implemented within DIRECT-INFO – is needed in order to fulfill these requirements (Rehatschek, 2004).

Within this context text analysis has been applied to documents reporting on entities, like football teams, that have close relations to large sponsoring companies. In the text analysis component of the system we had to detect if an entity was mentioned positively, negatively or neutrally. Besides all the processing and annotation issues to positive or negative mentions, we had to make our results available to a global MPEG-7 document, which is encoding the annotation results of various analysis of the modalities in-

volved (logo detection, speech recognition, text analysis etc.). This global MPEG-7 document was the input for a fusion component.

In the next sections we describe the Text Analysis (TA) component of DIRECT-INFO. We then briefly describe the linguistic description scheme (LDS) of MPEG-7 and show the annotation generated by the TA. Finally we briefly discuss the role the LDS, and generally speaking MPEG-7, can play in supporting an interoperable cross-media annotation strategy. It seems to us, that LDS is offering a good mean for adding semantic metadata to image/video, but not for a real semantic integration of text and media content annotation, which in the case of DIRECT-INFO was performed by an additional fusion component.

2 The detection of positive/negative mentioning

Our work in DIRECT-INFO has been dedicated in enhancing an already existing tool for linguistic annotation. This tool, called SCHUG (Shallow and CHunk-based Unification Grammar tool), is annotating texts considering both linguistic constituency and dependency structures (T. Declerck, M. Vela 2005).

A first development step was dedicated in creating specialized lexicons for various types of lexical categories (like nouns, adjectives and verbs) that can bear the property of being intrinsically positive or negative in a specific domain, as can be seen just below in the case of soccer:

```
command => {POS => Noun, INT => "positive"}  
dominate => {POS => Verb, INT => "positive"}  
weak => {POS => Adj, INT => "negative"}
```

Considering a sentence like “ManU takes the command in the game against the weak Spanish

team”, the head-noun of the direct object (linguistically speaking) “the command” gets from the access to the specialized DIRECT-INFO lexicon a tag “INTERPRETATION” with value “positive”. Whereas the adjective “weak” in the PP-adjunct “in the game against the weak Spanish team” gets an “INTERPRETATION” tag with value “negative”.

Once the words in the sentence have been lexically tagged with respect to their interpretation, the computing of the pos./neg. interpretation at the level of linguistic fragments and then at the level of the sentences can start. For this we have defined heuristics along the lines of the dependency structures delivered by the linguistic analysis. So in the case of the NP “the weak Spanish team”, the head noun “team”, as such a neutral expression, is getting the “INTERPRETATION” tag with the value “negative”, since it is modified by a “negative” adjective. In case the reference resolution algorithm of the linguistic tools has been able to specify that the “Spanish team” is in fact “Real Madrid” this entity gets a negative “INTERPRETATION” tag.

The head noun of the NP realizing the subject of the sentence, “ManU” gets a positive mention tag, since it is the subject of a positive verb and direct object combination (the NP “the command” having a positive reading, whereas the verb “takes” has a neutral reading).

A last aspect to be mentioned here concerns the treatment of the so-called polarity items. Specific words in natural language intrinsically carry a negation or position force (or scope). So the words *not*, *none* or *no* have an intrinsic negation force and negate the words and fragments in the context in which those specific words are occurring. The context that is negated by such words can be also called the “scope” (or the range) of the negation. Consider for example the sentence: “I would definitely pay £15 million to get Owen, *not even* a decent striker, instead...” Our tools are able to detect that the NP “decent striker” is negated, and therefore the positive reading of “decent striker” is being ruled out.

3 Metadata Description

The different content analysis modules of the DIRECT-INFO system extract different types of metadata, ranging from low-level audiovisual feature descriptions to semantic metadata. The global metadata description must be rich and has to clearly interrelate the various analysis results, as it is the input of the fusion component.

4.1 Using MPEG-7 for Detailed Description of Audiovisual Content

In DIRECT-INFO the MPEG-7 standard is used for metadata description. It is an excellent choice for describing audiovisual content because of its comprehensiveness and flexibility. The comprehensiveness results from the fact that the standard has been designed for a broad range of applications and thus employs very general and widely applicable concepts. The standard contains a large set of tools for diverse types of annotations on different semantic levels. The flexibility of MPEG-7, which is provided by a high level of generality, makes it usable for a broad application area without imposing strict constraints on the metadata models of these applications. The flexibility is very much based on the structuring tools and allows the description to be modular and on different levels of abstraction. MPEG-7 supports fine grained description, and it is possible to attach descriptors to arbitrary segments on any level of detail of the description.

Among the descriptive tools developed within the MPEG-7 framework, one is concerned with the use of natural language for adding metadata to the content description of image and video: the so-called Linguistic Description Scheme (LDS).

4.2 MPEG-7: The Linguistic Description Scheme (LDS)

MPEG-7 foresees four kinds of textual annotation that can be attached as metadata to some audio-video content. The natural language expression used here is “Spain scores a goal against Sweden. The scoring player is Morientes”.

Free Text Annotation: Here only tags are put around the text:

```
<TextAnnotation>
  <FreeTextAnnotation xml:lang="en">
    Spain scores a goal against Sweden.
    The scoring player is Morientes.
  </FreeTextAnnotation>
</TextAnnotation>
```

Key Word Annotation: Key Words are extracted from text and correspondingly annotated:

```
<TextAnnotation>
  <KeywordAnnotation>
    <Keyword>score</Keyword>
    <Keyword>Sweden</Keyword>
    <Keyword>Spain</Keyword>
    <Keyword>Morientes</Keyword>
  </KeywordAnnotation>
</TextAnnotation>
```

Structured Annotation: Question/Answering like semantics is associated to the text:

```
<TextAnnotation>
  <StructuredAnnotation>
    <Who><Name>Spain</Name></Who>
    <WhatAction><Name>score
goal</Name></WhatAction>
    <Where><Name>A Coruña,
Spain</Name></Where>
    <When><Name>March 25,
1998<Name></When>
  </StructuredAnnotation>
</TextAnnotation>
```

Dependency Structure: Here the full linguistic apparatus is used for annotating the text:

```
<TextAnnotation>
  <DependencyStructure>
    <Sentence>
      <Phrase operator="subject">
        <Head type="noun">Spain</Head>
      </Phrase>
      <Head type="verb" base-
Form="score">scored</Head>
      <Phrase operator="object">
        <Head type="article noun">a
goal</Head>
      </Phrase>
      <Phrase>
        <Head
type="preposition">against</Head>
      </Phrase>
      <Head>Sweden</Head></Phrase>
    </Sentence>
  </DependencyStructure>
</TextAnnotation>1
```

4 MPEG-7 Format of the Text Analysis component in DIRECT-INFO

On the base of the linguistic analysis of our dependency parser, we generate the “structured annotation” of the MPEG-7 Linguistic Description Scheme. We think that this kind of annotation is the most practical of LDS for adding semantics to multimedia content, since it is probably more intuitive for the media expert as the underlying linguistic dependency structure. At the same time it seems also straightforward to go first for a (internal) dependency analysis, since it is then relatively easy to map automatically dependency units to the “Who”, “WhatAction” and other tags of LDS.

The MPEG-7 output of the TA module of DIRECT-INFO looks like:

```
<MediaInformation>
  <MediaProfile>
    <MediaFormat>
      <Content href="http://www.direct-
info.net/mpeg7/cs/ContentCS.2004.xml/di.
content.writtenText">
        <Name>Written text</Name>
```

```
</Content>
  </MediaFormat>
  <MediaInstance>
    <InstanceIdentifier/>
    <MediaLocator>
      <!-- essence id-->
      <MediaUri>5543</MediaUri>
    </MediaLocator>
  </MediaInstance>
  </MediaProfile>
</MediaInformation>
<StructuralUnit href="http://www.direct-
info.net/mpeg7/cs/StructuralUnitCS.2004.
xml/di.vis.pdf">
  <Name>PDF</Name>
</StructuralUnit>
<!-- more than one page can be stored
within a file -->
<SpatialDecomposition criteria="Page">
  <StillRegion id="TA_PAGE1">
    <StructuralUnit
href="http://www.direct-
info.net/mpeg7/cs/StructuralUnitCS.2004.
xml/di.vis.page">
      <Name>Page</Name>
    </StructuralUnit>
    <SpatialDecomposition crite-
ria="TextAnalysis" gap="true" over-
lap="false">
      <StillRegion>
        <StructuralUnit
href="http://www.direct-
info.net/mpeg7/cs/StructuralUnitCS.2004.
xml/di.vis.textAnal ysisAnnotation">
          <Name>Text analysis annota-
tion</Name>
        </StructuralUnit>
        <TextAnnotation>
          <StructuredAnnotation>
            <WhatObject
href="http://www.direct-
info.net/mpeg7/cs/LogoCS.2004.xml/di.ta.
object.juventus">
              <Name
xml:lang="it">Juventus</Name>
            </WhatObject>
            <WhatAction
href="http://www.direct-
info.net/mpeg7/cs/TextAnalysisCS.2004.xml/
di.ta.action.teamMentioned">
              <Name xml:lang="it">mentioning of
team</Name>
            </WhatAction>
            <Why>
              <Name xml:lang="it">
295 771120 Con DVD Auto da Sogno Porsche
e 10, con calendario ufficiale 2006 Ju-
ventus o Milan" o Inter o Palermo o
Fiorentina o Totti" o Wrestling" e 6, 9
Euro 1, Poste Italiane Sped . in A.P
            </Name>
            </Why>
            <How href="http://www.direct-
info.net/mpeg7/cs/TextAnalysisCS.2004.xml/
di.ta.mentioning.neut">
              <Name xml:lang="it">neut</Name>
            </How>
          </StructuredAnnotation>
        </TextAnnotation>
      </StillRegion>
```

Without going into too much detail here, it is enough to stress that in the first part of the annotation, the link to the general multimedia and multimodal repository is ensured. We have to

¹ These examples are taken from a former and excellent online tutorial on MPEG-7 by Philippe Salembier.

deal with a PDF document that should be processed by a Text Analysis tool. The “essence” ID is giving information about the location where the application relevant data is stored and where the results of the Text Analysis should be stored. All this metadata is ensuring the combination of the results of the analysis of various modalities dealing with one application relevant dataset (for example the combination of the logo detection of a brand and the related positive or negative mentioning of a team sponsored by this brand). For reason of place, we can not show and comment here the complete (and multimodal) MPEG-7 annotation, but details are given in (G. Kienast, 2005).

The second part of the annotation gives the results of the combined linguistic and “structured” analysis we are dealing with. As mentioned above, in the case of DIRECT-INFO, results of text analysis are accessed via the structured annotation of the Linguistic Description Schema of MPEG-7.

5 Conclusions and future Work

In the DIRECT-INFO project we managed to include results of text analysis in an automated fashion into a MPEG-7 description, which was dealing with the XML representation of the analysis of various modalities. Using corresponding metadata, it was possible to ensure the encoding/annotation of the related results in one file and to facilitate the access to the separated annotation using XPath. As such the DIRECT-INFO MPEG-7 annotation schema is offering a practicable multi-dimensional annotation scheme, if we consider a “dimensions” as being the output of the analysis of various modalities. MPEG-7 proved to be generic and flexible enough for combining, saving and accessing various types of annotation.

Limitations of MPEG-7 were encountered when the task was about fusion or merging of information encoded in the various descriptors (or features), and this task was addressed in a posterior step, whereas the encoding scheme of MPEG-7 was not longer helpful, in defining for example relations between the annotation resulting from the different modules or for defining constraints between those annotation. There seems to be a need for a higher level of representation for annotation resulting from the analysis of distinct media, being low-level features for images or high-level semantic features for texts.

The need of an “ontologization” of multimedia features has been already recognized and projects are already dealing with this, like AceMedia. Initial work in relating multimodal annotation in DIRECT-INFO will be further developed in K-Space, a new Network of Excellence, which goal is to provide for support in semantic inference for both automatic and semi-automatic annotation and retrieval of multimedia content. K-Space aims at closing the “semantic gap” between the low-level content descriptions and the richness and subjectivity of semantics in high-level human interpretations of audiovisual media.

6 Acknowledgements

The R&D work presented in this paper was partially conducted within the DIRECT-INFO project, funded under the 6th Framework Programme of the European Community within the strategic objective "Semantic-based knowledge management systems" (IST FP6-506898). Actual work on interoperability of media, language and semantic annotation is being funded by the Network of Excellence K-Space (IST FP6-027026).

References

- T. Declerck, J. Kuper, H. Saggion, A. Samiotou, P. Wittenburg, J. Contreras. Contribution of NLP to the Content Indexing of Multimedia Documents. In *Lecture Notes in Computer Science Volume 3115 / 2004* Pages 610-618, Springer-Verlag Heidelberg, 6 2004.
 - T. Declerck, M. Vela, “Linguistic Dependencies as a Basis for the Extraction of Semantic Relations”, in Proceedings of the ECCB'05 Workshop on Biomedical Ontologies and Text Processing, Madrid (2005)
 - G. Kienast, A. Horti, András, H. Rehatschek, S. Busemann, T. Declerck, V. Hahn and R. Cavet. “DIRECT INFO: A Media Monitoring System for Sponsorship Tracking.” In *Proceedings of the ACM SIGIR Workshop on Multimedia Information Retrieval*. 2005.
 - H. Rehatschek: "DIRECT-INFO: Media monitoring and multimodal analysis for time critical decisions". Proceedings of the 5th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), ISBN-972-98115-7-1, Lisbon, April 2004.
- AceMedia project: <http://www.acemedia.org/aceMedia>
DIRECT-INFO project: <http://www.direct-info.net/>
K-Space project: <http://kspace.qmul.net/>
MPEG-7: <http://www.chiariglione.org/mpeg/>