A Linked Data Model for Multimodal Sentiment and Emotion Analysis

J. Fernando Sánchez-Rada
Grupo de Sistemas
Inteligentes
Universidad Politécnica
de Madrid
`jfernando@dit.upm.es`

Carlos A. Iglesias
Grupo de Sistemas
Inteligentes
Universidad Politécnica
de Madrid
`cif@dit.upm.es`

Ronald Gil
Massachusetts Institute of
Technology
`rongil@mit.edu`

Abstract

The number of tools and services for sentiment analysis is increasing rapidly. Unfortunately, the lack of standard formats hinders interoperability. To tackle this problem, previous works propose the use of the NLP Interchange Format (NIF) as both a common semantic format and an API for textual sentiment analysis. However, that approach creates a gap between textual and sentiment analysis that hampers multimodality. This paper presents a multimedia extension of NIF that can be leveraged for multimodal applications. The application of this extended model is illustrated with a service that annotates online videos with their sentiment and the use of SPARQL to retrieve results for different modes.

1 Introduction

With the rise of social media and crowdsourcing, the interest in automatic means of extraction and aggregation of user opinions (Opinion Mining) and emotions (Emotion Mining) is growing. This tendency is mainly focused on text analysis, the cause and consequence of this being that the tools for text analysis are getting better and more accurate. As is often the case, these tools are heterogeneous and implement different formats and APIs. This problem is hardly new or limited to sentiment analysis, it is also present in the Natural Language Processing (NLP) field. In fact, both fields are closely related: textual sentiment analysis can be considered a branch of NLP. Looking at how NLP deals with heterogeneity and interoperability we find NIF, a format for NLP services

that solves these issues. Unfortunately, *NLP Interchange Format* (NIF) (Hellmann et al., 2013) is not enough to annotate sentiment analysis services. Fortunately, it can be extended, by exploiting the extensibility of semantic formats. Using this extensibility and already existing ontologies for the sentiment and emotion domains, the R&D Eurosentiment project recently released a model that extends NIF for sentiment analysis (Buitelaar et al., 2013).

However, the Eurosentiment model is bound to textual sentiment analysis, as NIF focuses on annotation of text. The R&D MixedEmotions project aims at bridging this gap by providing a Big Linked Data Platform for multimedia and multilingual sentiment and emotion analysis. Hence, different modes (e.g. images, video, audio) require different formats. Format heterogeneity becomes problematic when different modes coexist or when the text is part of other media. Some examples of this include working with text extracted from a picture with OCR, or subtitles and transcripts of audio and video. This scenario is not uncommon, given the maturity of textual sentiment analysis tools.

In particular, this paper focuses on video and audio sources that contain emotions and opinions, such as public speeches. We aim to represent that information in a linked data format, linking the original source with its transcription and any sentiments or emotions found in any of its modes. Using the new model it is possible to represent and process multimodal sentiment information using a common set of tools.

The rest of the paper is structured as follows: Section 2 covers the background for this work; Section 3 presents requirements for semantic annotation of sentiment in multimedia; Section 4

introduces the bases for sentiment analysis using NIF and delves into the use of NIF for media other than text; Section 5 exemplifies the actual application of the new model with a prototype and semantic queries; Section 6 is dedicated to related work; lastly, Section 7 summarises the conclusions drawn from our work and presents possible lines of work.

2 Background

2.1 Annotation based on linked data

Annotating is the process of associating metadata with multimedia assets. Previous research has shown that annotations can benefit from compatibility with linked data technologies (Hausenblas, 2007).

The W3C Open Annotation Community Group has worked towards a common RDF-based specification for annotating digital resources. The group intends to reconcile two previous proposals: the Annotation Ontology (Ciccarese et al., 2011) and the Open Annotation Collaboration (OAC) (Haslhofer et al., 2011). Both proposals incorporate elements from the earlier Annotea model (Kahan et al., 2002). The Open Annotation Ontology (Robert Sanderson and de Sompel, 2013) provides a general description mechanism for sharing annotation between systems based on an RDF model. An annotation is defined by two relationships: *body*, the annotation itself, and *target*, the asset that is annotated. Both body and target can be of any media type. In addition, parts of the body or target can be identified by using Fragment Selectors (oa:FragmentSelector) entities. W3C Fragment URIs (Tennison, 2012) can be used instead, although the use of Fragment Selectors is encouraged. The vocabulary defines fragment selectors for text (oa:Text), text segments plus passages before or after them (oa:TextQuoteSelector), byte streams (oa:DataPositionSelector), areas (oa:AreaSelector), states (oa:State), time moments (oa:TimeState) and request headers (oa:RequestHeaderState). Finally, Open Annotation (OA) ontology defines how annotations are published and transferred between systems. The recommended serialisation format is JSON-LD.

Another research topic has been the standardisation of linguistic annotations in order to improve the interoperability of NLP tools and resources. The main proposals are Linguistic Annotation Framework (LAF) and NIF 2.0. The ISO Specification LAF (Ide and Romary, 2004) and its extension Graph Annotation Format (GrAF) (Ide and Suderman, 2007) define XML serialisation of linguistic annotation as well as RDF mappings. NIF 2.0 (Hellmann et al., 2013) follows a pragmatic approach to linguistic annotations and is focused on interoperability of NLP tools and services. It is directly based on RDF, Linked Data and ontologies, and it allows handling structural interoperability of linguistic annotations as well as semantic interoperability. NIF 2.0 Core ontology provides classes and properties to describe the relationships between substrings, text and documents by assigning URIs to strings. These URIs can then be used as subjects in RDF easily annotated. NIF builds on current best practices for counting strings and creating offsets such as LAF. NIF uses Ontologies for Linguistic Annotation (OLiA) (Chiarcos, 2012) to provide stable identifiers for morpho-syntactical annotation tag sets. In addition to the core ontology, NIF defines Vocabulary modules as an extension mechanism to achieve interoperability between different annotation layers. Some of the defined vocabularies are Marl (Westerski et al., 2011) and Lexicon Model for Ontologies (lemon) (Buitelaar et al., 2011).

As discussed by Hellmann (Hellmann, 2013), the selection of the annotation scheme comes from the domain annotation requirements and the trade-off among granularity, expressiveness and simplicity. He defines different profiles with this purpose. The profile NIF simple can express *the best estimate* of an NLP tool in a flat data model, with a low number of triples. An intermediate profile called NIF Stanbol allows the inclusion of alternative annotations with different confidence as well as provenance information that can be attached to the additionally created URN for each annotation. This profile is integrated with the semantic content management system Stanbol (Westenhaler, 2014). Finally, the profile NIF OA provides the most expressive model but requires more triples and creates up to four new URNs per annotation, making it more difficult to query.

Finally, we review Fusepool since they propose an annotation model that combines OA and NIF. Fusepool (Westenhaler, 2014) is an R&D project whose purpose is to digest and turn data from different sources into linked data to make data interoperable for reuse. One of the tasks of this

project is to define a new Enhancement Structure for the semantic content management system Apache Stanbol (Bachmann-Gmür, 2013). Fusepool researchers' main design considerations with OA is for it to define a very expressive model capable of very complex annotations. This technique comes with the disadvantage of needing a high amount of triples to represent lower level NLP processing, which in turn complicates the queries necessary to retrieve simple data.

2.2 Eurosentiment Model

The work presented here is partly based on an earlier work (Buitelaar et al., 2013) developed within the Eurosentiment project. The Eurosentiment model proposes a linked data approach for sentiment and emotion analysis, and it is based on the following specifications:

- Marl (Westerski et al., 2011) is a vocabulary designed to annotate and describe subjective opinions expressed on the web or in information systems

- Onyx (Sanchez-Rada and Iglesias, 2013) is built on the same principles as Marl to annotate and describe emotions, and provides interoperability with Emotion Markup Language (EmotionML) (Schröder et al., 2011)

- lemon (Buitelaar et al., 2011) defines a lexicon model based on linked data principles which has been extended with Marl and Onyx for sentiment and emotion annotation of lexical entries

- NIF 2.0 (Hellmann et al., 2013) which defines a semantic format and API for improving interoperability among natural language processing services

The way these vocabularies have been integrated is illustrated in the example below, where we are going to analyse the sentiment of an opinion ("Like many Paris hotels, the rooms are too small") posted in TripAdvisor. In the Eurosentiment model, *lemon* is used to define the lexicon for a domain and a language. In our example, we have to generate this lexicon for the hotel domain and the English language[1]. A reduced lexicon for Hotels in English (le:hotel_en) is shown in Listing 1

---

[1]The reader interested in how this domain specific lexicon can be generated can consult (Vulcu et al., 2014).

for illustration purposes. The lexicon is composed of a set of lexical entries (prefix lee). Each lexical entry is semantically disambiguated and provides a reference to the syntactic variant (in the example the canonical form) and the senses. The example shows how the senses have been extended to include sentiment features. In particular, the sense small_1 in the context of room_1 has associated a negative sentiment. That is, "small room" is negative (while small phone could be positive, for example).

```
lee:sense/small_1 a lemon:Sense;
 lemon:reference "01391351";
 lexinfo:partOfSpeech lexinfo:adjective;
 lemon:context lee:sense/room_1;
 marl:polarityValue "-0.5"^^xsd:double;
 marl:hasPolarity marl:negative.

le:hotel_en a lemon:Lexicon;
 lemon:language "en";
 lemon:topic ed:hotel;
 lemon:entry lee:room, lee:Paris, lee:
     small.

lee:room a lemon:LexicalEntry;
 lemon:canonicalForm [ lemon:writtenRep
     "room"@en ];
 lemon:sense [ lemon:reference wn:synset
     -room-noun-1;
     lemon:reference dbp:Room ];
 lexinfo:partOfSpeech lexinfo:noun.

lee:Paris a lemon:LexicalEntry;
  lemon:canonicalForm [ lemon:writtenRep
      "Paris"@en ];
  lemon:sense [ lemon:reference dbp:
      Paris;
      lemon:reference wn:synset-room-noun
         -1 ];
 lexinfo:partOfSpeech lexinfo:noun.

lee:small a lemon:LexicalEntry;
 lemon:canonicalForm [ lemon:writtenRep
     "small"@en ];
 lemon:sense lee:sense/small_1;
 lexinfo:partOfSpeech lexinfo:adjective.
```

Listing 1: Sentiment analysis expressed with Eurosentiment model.

The Eurosentiment model uses NIF in combination with Marl and Onyx to provide a standardised service interface. In our example, let us assume the opinion has been published at http://tripadvisor.com/myhotel. NIF follows a linked data principled approach so that different tools or services can annotate a text. To this end, texts are converted to RDF literals and an URI is generated so that annotations can be defined for that text in a linked data way. NIF offers different URI Schemes to identify text fragments inside a

string, e.g. a scheme based on RFC5147 (Wilde and Duerst, 2008), and a custom scheme based on context. In addition to the format itself, NIF 2.0 defines a REST API for NLP services with standardised parameters. An example of how these ontologies are integrated is illustrated in Listings 2, 3 and 4.

```
<http://tripadvisor.com/myhotel#char
    =0,49>
  rdf:type nif:RDF5147String , nif:
      Context;
  nif:beginIndex "0";
  nif:endIndex "49";
  nif:sourceURL <http://tripadvisor.com/
      myhotel.txt>;
  nif:isString "Like many Paris hotels,
      the rooms are too small";
  marl:hasOpinion <http://tripadvisor.
      com/myhotel/opinion/1>.
```

Listing 2: NIF + Marl output of a service call http://eurosentiment.eu?i=Like many Paris hotels, the rooms are too small

```
<http://tripadvisor.com/myhotel/opinion
    /1>
  rdf:type marl:Opinion;
  marl:describesObject dbp:Hotel;
  marl:describesObjectPart dbp:Room;
  marl:describesFeature  "size";
  marl:polarityValue "-0.5";
  marl:hasPolarity: http://purl.org/marl
      /ns#Negative.
```

Listing 3: Sentiment analysis expressed with Eurosentiment model.

```
<http://eurosentiment.eu/analysis/1>
  rdf:type marl:SentimentAnalysis;
  marl:maxPolarityValue "1";
  marl:minPolarityValue "-1";
  marl:algorithm "dictionary-based";
  prov:used le:hotel_en;
  prov:wasAssociatedWith http://dbpedia.
      org/resource/UPM.
```

Listing 4: Sentiment analysis expressed with Eurosentiment model.

3 Requirements for semantic annotation of sentiment in multimedia resources

The increasing need to deal with human factors, including emotions, on the web has led to the development of the W3C specification EmotionML (Schröder et al., 2011). EmotionML aims for a trade-off between practical applicability and scientific well-foundedness. Given the lack of agreement on a finite set of emotion descriptors,

EmotionML follows a plug-in model where emotion vocabularies can be defined depending on the application domain and the aspect of emotions to be focused.

EmotionML (Schröder et al., 2011) uses Media URIs to annotate multimedia assets. Temporal clipping can be specified either as Normal Play Time (npt) (Schulzrinne et al., 1998), as SMPTE timecodes (Society of Motion Picture and Television Engineers, 2009), or as real-world clock time (Schulzrinne et al., 1998).

During the definition of the EmotionML specification, the Emotion Incubator group defined 39 individual use cases (Schröder et al., 2007) that could be grouped into three broad types: manual annotation of materials (e.g. annotation of videos, speech recordings, faces or texts), automatic recognition of emotions from sensors and generation of emotion-related system responses. Based on these uses cases as well as others identified in the literature (Grassi et al., 2011), a number of requirements have been identified for the annotation of multimedia assets based on linked data technologies:

- **Standards compliance**. Emotion annotations should be based on linked data technologies such as RDF or W3C Media Fragment URI. Unfortunately, EmotionML has been defined in XML. Nevertheless, as commented above, the vocabulary Onyx provides a linked data version of EmotionML that can be used instead. Regarding the annotation framework, OA covers the annotation of multimedia assets while NIF only supports the annotation of textual sources.

- **Trace annotation of time-varying signals**. The time curve of properties scales (e.g. arousal or valence) should be preserved. To this end, EmotionML defines two mechanisms. The element *trace* allows the representation of the time evolution of a dynamic scale value based on a periodic sampling of values (i.e. one value every 100ms at 10 Hz). In case of aperiodic sampling, separate emotion annotations should be used. The current version of the ontologies we use does not support trace annotations.

- **Annotations of multimedia fragments**. Fragments of multimedia assets should be enabled. To this end, EmotionML uses Media

URIs to be able to annotate temporal interval or frames. As presented above, NIF provides a compact scheme for textual fragment annotation, but it does not cover multimedia fragments. In contrast, OA supports the annotation of multimedia fragments using a number of triples.

- **Collaborative and multi-modal annotations**. Emotion analysis of multimedia assets may be performed based on different combination of modalities (i.e. full body video, facial video, each with or without speech or textual transcription). Thus, interoperability of emotion annotations is essential. Semantic web technologies provide a solid base for distributed, interoperable and shareable annotations, with proposals such as OA and NIF.

4 Linked Data Annotation for Multimedia Sentiment Analysis

One of the main goals of NIF is interoperability between NLP tools. For this, it uses a convention to assign URIs to parts of a text. Since URIs are unique, different tools can analyse the same text independently, and one may use the URIs later to combine the information from both.

These URIs are constructed with a combination of the URI of the source of the string (its context), and a unique identifier for that string within that particular context. A way to assign that identifier is called a URI scheme. Strings belong to different classes, according to the scheme used to generate its URI. The currently available schemes are: ContextHashBasedString, OffsetBasedString, RFC5147String and ArbitraryString. The usual scheme is *RFC5147String*.

For instance, for a context `http://example.com`, its content may be "This is a test", and the *RFC5147String* `http://example.com#char=5,7` would refer to the "is" part within the context.

However, to annotate multimedia sources indexing by characters is obviously not possible. We need a different way to uniquely refer to a fragment.

Among the different possible approaches to identify media elements, we propose to follow the same path as the Ontology for Media Resources (Lee et al., 2012) and use the Media Fragments URI W3C recommendation (Troncy et al.,

2012). The recommendation specifies how to refer to a specific fragment or subpart of a media resource. URIs follow this scheme:

`<scheme>:<part>[?<q>][#<frag.>]`

Where `<scheme>` is the specific scheme or protocol (e.g. http), `part` is the hierarchical part (e.g. example.com), `q` is the query (e.g. user=Nobody), and `frag` is the piece we are interested in: the multimedia fragment (e.g. t=10).

Since the Media Fragments URI schema is very similar to those already used in NIF and follows the same philosophy, we have extended NIF to include it. The result is Figure 1.
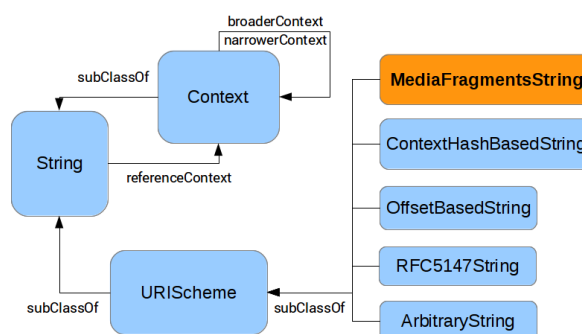

Figure 1: By extending the URI Schemes of NIF, we make it possible to use multimedia sources in NIF, and refer to their origin using the Media Fragments recommendation.

Using this URI Scheme and the NIF notation for sentiment analysis, the results from a service that analyses both the captions from a YouTube video and the video comments would look like the document in Listing 5. In this way, we fulfill the requirements previously identified in Sect. 3. This example is, in fact, the aggregation of three different kinds of analysis: textual sentiment analysis on comments (`CommentAnalysis`) and captions (`CaptionAnalysis`), and sentiment analysis based on facial expressions (`SmileAnalysis`). Each analysis would individually return a document similar to that of the example, with only the fields corresponding to that particular analysis.

The results can be summarised as follows: a youtube video (`http://youtu.be/W07PoKUD-Yk`) is tagged as positive overall based on facial expressions (`OpinionS01`); the section of the video from second 108 to second 110 (`http://youtu.be/W07PoKUD-Yk#t=108, 110`) reflects negative sentiment judging by the captions (`OpinionT01`);

lastly, the video has a comment (`http://www.youtube.com/comment?lc=<CommentID>`) that reflects a positive opinion (`OpinionC01`).

The JSON-LD context in Listing 6 provides extra information the semantics of the document, and has been added for completeness.

```
{
"analysis": [{
 "@id": "SmileAnalysis",
 "@type": "marl:SentimentAnalysis",
 "marl:algorithm": "AverageSmiles"
 }, {
 "@id": "CaptionAnalysis",
 "@type": "marl:SentimentAnalysis",
 "marl:algorithm": "NaiveBayes"
 }, {
 "@id": "CommentAnalysis",
 "@type": "marl:SentimentAnalysis",
 "marl:algorithm": "NaiveBayes"
 }],
"entries": [{
 "@id": "http://youtu.be/W07PoKUD-Yk",
 "@type": [
  "nifmedia:MediaFragmentsString",
  "nif:Context"],
 "nif:isString": "<FULL Transcript>",
 "opinions": [{
 "@id": "_:OpinionS01",
 "marl:hasPolarity": "marl:Positive",
 "marl:polarityValue": 0.5,
 "prov:generatedBy": "SmileAnalysis"
 }],
 "sioc:hasReply": "http://
   ↪ www.youtube.com/comment?lc=<
   ↪ CommentID>",
 "strings": [{
 "@id": "http://youtu.be/W07PoKUD-Yk#t=
   ↪ 108,110",
 "@type": "nifmedia:
   ↪ MediaFragmentsString",
 "nif:anchorOf": "Family budgets under
   ↪ pressure",
 "opinions": [{
  "@id": "_:OpinionT01",
  "marl:hasPolarity": "marl:Negative",
  "marl:polarityValue": -0.3058,
  "prov:generatedBy": "CaptionAnalysis"
  }]
 }]
 }, {
 "@id": "http://www.youtube.com/comment?
   ↪ lc=<CommentID>",
 "@type": [
  "nif:Context", "nif:RFC5147String" ],
 "nif:isString": "He is well spoken",
 "opinions": [{
 "@id": "OpinionC01",
 "marl:hasPolarity": "marl:Positive",
 "marl:polarityValue": 1,
 "prov:generatedBy": "CommentAnalysis"
 }]
 }]
}
```

Listing 5: Service results are annotated on the fragment level with sentiment and any other property in NIF such as POS tags or entities.

```
{
 "marl": "http://www.gsi.dit.upm.es/
    ↪ ontologies/marl/ns#",
 "nif": "http://persistence.uni-
    ↪ leipzig.org/nlp2rdf/ontologies/
    ↪ nif-core#",
 "onyx": "http://www.gsi.dit.upm.es/
    ↪ ontologies/onyx/ns#",
 "nifmedia": "http://www.gsi.dit.upm.es/
    ↪ ontologies/nif/ns#",
 "analysis": {
    "@id": "prov:wasInformedBy"
 },
 "opinions": {
    "@container": "@list",
    "@id": "marl:hasOpinion",
    "@type": "marl:Opinion"
 },
 "entries": {
    "@id": "prov:generated"
 },
 "strings": {
    "@reverse": "nif:hasContext"
 }
}
```

Listing 6: JSON-LD context for the results necessary to give semantic meaning to the JSON in Listing 5.

5 Application

5.1 VESA: Online HTML5 Video Annotator

The first application to use NIF annotation for sentiment analysis of Multimedia sources is VESA, the Video Emotion and Sentiment Analysis tool. VESA is both a tool to run sentiment analysis of online videos, and a visualisation tool which shows the evolution of sentiment information and the transcript as the video is playing, using HTML5 widgets. The visualisation tool can run the analysis in real time (live analysis) or use previously stored results.

The live analysis generates transcriptions using the built-in Web Speech API in Google Chrome[2] while the video plays in the background. To improve the performance and accuracy of the transcription process, the audio is chunked in sentences (delimited by a silence). Then, each chunk is sent to a sentiment analysis service. As of this writing, users can choose sentiment analysis in Spanish or English, in a general or a financial domain, using different dictionaries.

The evolution of sentiment within the video is shown as a graph below the video in Figure 2. The

---

[2] `https://www.google.com/intl/en/chrome/demos/speech.html`

full transcript of the video allows users to check the accuracy of the transcription service.

The results from the service can be stored in a database, and can be later replayed. We also developed a Widget version of the annotator that can be embedded in other websites, and integrated in widget frameworks like Sefarad[3].

The project is completely open source and can be downloaded from its Github repository[4].
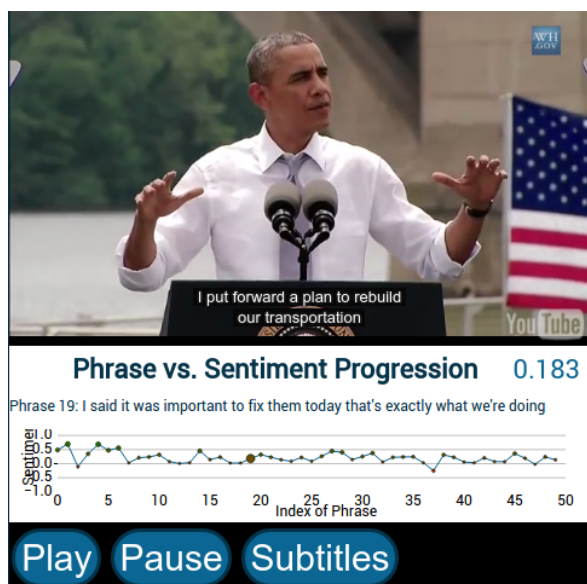

Figure 2: The graph shows the detected sentiment in the video over time, while the video keeps playing.

5.2 Semantic multimodal queries

This section demonstrates how it would be possible to integrate sentiment analysis of different modes using SPARQL. In particular, it covers two scenarios: fusion of results from different modes, and detection of complex patterns using information from several modes.

As discussed in Section 6, SPARQL has some limitations when it comes to querying media fragments. There are extensions to SPARQL that overcome those limitations. However, for the sake of clarity, this section will avoid those extensions. Instead, the examples assume that the original media is chunked equally for every mode. Every chunk represents a media fragment, which may contain an opinion.

When different modes yield different sentiments or emotions, it is usually desirable to integrate all the results into a single one. The query in Listing 7 shows how to retrieve all the opinions for each chunk. These results can be fed to a fusion algorithm.

```
SELECT ?frag ?algo ?opinion ?pol WHERE {
  ?frag a nifmedia:MediaFragmentsString;
       marl:hasOpinion ?opinion.
  ?opinion marl:hasPolarity ?pol.
  ?algo prov:generated ?opinion.
}
```

Listing 7: Gathering all the opinions detected in a video.

Another possibility is that the discrepancies between different modes reveal useful information. For instance, using a cheerful tone of voice for a negative text may indicate sarcasm or untruthfulness. Listing 8 shows an example of how to detect such discrepancies. Note that it uses both opinions and emotions at the same time.

```
SELECT ?frag WHERE {
  ?frag a nifmedia:MediaFragmentsString;
       marl:hasOpinion ?opinion;
       onyx:hasEmotion ?emo.
  ?opinion prov:wasGeneratedBy
            _:TextAnalysis;
         marl:hasPolarity marl:
            Negative.
  ?emo prov:wasGeneratedBy
            _:AudioAnalysis;
       onyx:hasEmotionCategory wna:
         Cheerfulness.
}
```

Listing 8: Detecting negative text narrated with a cheerful tone of voice.

6 Related work

Semedi research group proposes the use of semantic web technologies for video fragment annotation (Morbidoni et al., 2011) and affective states based on the HEO (Grassi et al., 2011) ontology. They propose the use of standards, such as XPointer (Paul Grosso and Walsh, 2003) and Media Fragment URI (Troncy et al., 2012) for defining URIs for text and multimedia, respectively, as well as the Open Annotation Ontology (Robert Sanderson and de Sompel, 2013) for expressing the annotations. Their approach is similar to the one we have proposed, based on web standards and linked data to express emotion annotations. Our proposal has been aligned with the latest available specifications, which have been extended as presented in this article.

On the other hand, a better integration between multimedia and the linked data toolbox would be necessary. Working with multimedia fragments in plain SPARQL is not an easy task. More specifically, it is the relationship between fragments that complicates it, e.g. finding overlaps or contiguous segments. An extension to SPARQL by Kurz et al. (Kurz et al., 2014), SPARQL-MM, introduces convenient methods that allow these operations in a concise way.

7 Conclusions and future work

We have introduced the conceptual tools to describe sentiment and emotion analysis results in a semantic format, not only from textual sources but also multimedia.

Despite being primarily oriented towards analysis of texts extracted from multimedia sources, this approach can be used to apply other kinds of analysis, in a way similar to how NIF integrates results from different tools. However, more effort needs to be put into exploring different use cases and how they can be integrated in our extension of NIF for sentiment analysis in multimedia. This work will be done in the project MixedEmotions, where several use cases (Brand Monitoring, Social TV or Call Center Management) have been identified and involve multimedia analysis.

In addition, this discussion can be carried out in the *Linked Data Models for Emotion and Sentiment Analysis* W3C Community Group [5], where professionals and academics of the Semantic and sentiment analysis worlds meet and discuss the application of an interdisciplinary approach.

Regarding the video annotator, although the current version is fully functional, it could be improved in several ways. The main limitation is that its live analysis relies on the Web Speech API, and needs user interaction to set specific audio settings. We are studying other fully client-side approaches.

Acknowledgements

The authors also want to thank Berto Yáñez for his support and inspiring demo "Popcorn.js Sentiment Tracker".

References

Reto Bachmann-Gmür. 2013. *Instant Apache Stanbol*. Packt Publisher.

Paul Buitelaar, Philipp Cimiano, John McCrae, Elena Montiel-Ponsoda, and Thierry Declerck. 2011. Ontology lexicalisation: The lemon perspective. In *Workshop at 9th International Conference on Terminology and Artificial Intelligence (TIA 2011)*, pages 33–36.

Paul Buitelaar, Mihael Arcan, Carlos A Iglesias, J Fernando Sánchez-Rada, and Carlo Strapparava. 2013. Linguistic linked data for sentiment analysis. In *2nd Workshop on Linked Data in Linguistics (LDL-2013): Representing and linking lexicons, terminologies and other language data*, Pisa, Italy.

Christian Chiarcos. 2012. Ontologies of linguistic annotation: Survey and perspectives. In *LREC*, pages 303–310.

Paolo Ciccarese, Marco Ocana, Leyla Jael Garcia-Castro, Sudeshna Das, and Tim Clark. 2011. An open annotation ontology for science on web 3.0. *J. Biomedical Semantics*, 2(S-2):S4.

Marco Grassi, Christian Morbidoni, and Francesco Piazza. 2011. Towards semantic multimodal video annotation. In *Toward Autonomous, Adaptive, and Context-Aware Multimodal Interfaces. Theoretical and Practical Issues*, volume 6456 of *Lecture Notes in Computer Science*, pages 305–316. Springer Berlin Heidelberg.

Bernhard Haslhofer, Rainer Simon, Robert Sanderson, and Herbert Van de Sompel. 2011. The open annotation collaboration (oac) model. In *Multimedia on the Web (MMWeb), 2011 Workshop on*, pages 5–9. IEEE.

Michael Hausenblas. 2007. Multimedia vocabularies on the semantic web. Technical report, World Wide Web (W3C).

Sebastian Hellmann, Jens Lehmann, Sören Auer, and Martin Brümmer. 2013. Integrating nlp using linked data. In *The Semantic Web–ISWC 2013*, pages 98–113. Springer.

Sebastian Hellmann. 2013. *Integrating Natural Language Processing (NLP) and Language Resources using Linked Data*. Ph.D. thesis, Universität Leipzig.

Nancy Ide and Laurent Romary. 2004. International standard for a linguistic annotation framework. *Natural language engineering*, 10(3-4):211–225.

---

[5] http://www.w3.org/community/sentiment/

Nancy Ide and Keith Suderman. 2007. Graf: A graph-based format for linguistic annotations. In *Proceedings of the Linguistic Annotation Workshop*, LAW '07, pages 1–8, Stroudsburg, PA, USA. Association for Computational Linguistics.

J. Kahan, M.-R. Koivunen, E. Prud'Hommeaux, and R.R. Swick. 2002. Annotea: an open {RDF} infrastructure for shared web annotations. *Computer Networks*, 39(5):589 – 608.

Thomas Kurz, Sebastian Schaffert, Kai Schlegel, Florian Stegmaier, and Harald Kosch. 2014. Sparql-mm-extending sparql to media fragments. In *The Semantic Web: ESWC 2014 Satellite Events*, pages 236–240. Springer.

WonSuk Lee, Werner Bailer, Tobias Bürger, Pierre-Antoine Champin, Jean-Pierre Evain, Véronique Malaisé, Thierry Michel, Felix Sasaki, Joakim Söderberg, Florian Stegmaier, and John Strassner. 2012. Ontology for Media Resources 1.0. Technical report, World Wide Web Consortium (W3C), February. Available at http://www.w3.org/TR/mediaont-10/.

C. Morbidoni, M. Grassi, M. Nucci, S. Fonda, and G. Ledda. 2011. Introducing semlib project: semantic web tools for digital libraries. *International Workshop on Semantic Digital Archives-Sustainable Long-term Curation Perspectives of Cultural Heritage Held as Part of the 15th International Conference on Theory and Practice of Digital Libraries (TPDL), Berlin*.

Jonathan Marsh Paul Grosso, Eve Mater and Normal Walsh. 2003. W3C XPointer Framework. Technical report, World Wide Web Consortium (W3C), March. Available at http://www.w3.org/TR/xptr-framework/.

Pablo Ciccarese Robert Sanderson and Herbert Van de Sompel. 2013. W3C Open Annotation Data Model. Technical report, World Wide Web Consortium (W3C), February. Available at http://www.openannotation.org/spec/core/.

J Fernando Sanchez-Rada and Carlos Angel Iglesias. 2013. Onyx: Describing emotions on the web of data. In *ESSEM@ AI* IA*, pages 71–82. Citeseer.

Marc Schröder, Enrico Zovato, Hannes Pirker, Christian Peter, and Felix Burkhardt. 2007. W3C Emotion Incubator Group Report 10 July 2007. Technical report, W3C, July. Available at http://www.w3.org/2005/Incubator/emotion/XGR-emotion/.

Marc Schröder, Paolo Baggia, Felix Burkhardt, Catherine Pelachaud, Christian Peter, and Enrico Zovato. 2011. Emotionml – an upcoming standard for representing emotions and related states. In Sidney D'Mello, Arthur Graesser, Björn Schuller, and Jean-Claude Martin, editors, *Affective Computing and Intelligent Interaction*, volume 6974 of *Lecture Notes in Computer Science*, pages 316–325. Springer Berlin Heidelberg.

H. Schulzrinne, A. Rao, and R. Lanphier. 1998. Real Time Streaming Protocol (RTP). Technical Report RFC2326, IETF. Available at http://www.ietf.org/rfc/rfc2326.txt.

Society of Motion Picture and Television Engineers. 2009. SMPTE - RP 136. time and control codes for 24, 25 or 30 frame-per-second motion-picture systems - stabilized 2009. Technical report, SMPTE.

Jeni Tennison. 2012. Best practices for fragment identifiers and media type definitions. Technical report, World Wide Web (W3C).

Raphaël Troncy, Erik Mannens, Silvia Pfeiffer, and Davy Van Deursen. 2012. W3C Recommendation Media Fragments URI 1.0. Technical report, World Wide Web Consortium (W3C), September. Available at http://www.w3.org/TR/2012/REC-media-frags-20120925/.

Gabriela Vulcu, Paul Buitelaar, Sapna Negi, Bianca Pereira, Mihael Arcan, Barry Coughlan, J. Fernando Sánchez-Rada, and Carlos A. Iglesias. 2014. Generating Linked-Data based Domain-Specific Sentiment Lexicons from Legacy Language and Semantic Resources. In *th International Workshop on Emotion, Social Signals, Sentiment & Linked Open Data, co-located with LREC 2014*, Reykjavik, Iceland, May. LREC2014.

Rupert Westenhaler. 2014. Open Annotation and NIF 2.0 based Annotation Model for Apache Stanbol. In *Proceedings of ISWC 2014*.

Adam Westerski, Carlos A. Iglesias, and Fernando Tapia. 2011. Linked opinions: Describing sentiments on the structured web of data.

E. Wilde and M. Duerst. 2008. URI Fragment Identifiers for the text/plain Media Type. Technical Report RDF5147, Internet Engineering Task Force (IETF), April. Available at https://tools.ietf.org/html/rfc5147.