I am a Machine, let me understand Web Media!

Magnus Knuth, Jörg Waitelonis, and Harald Sack

Hasso Plattner Institute, University of Potsdam, Germany,
{magnus.knuth|joerg.waitelonis|harald.sack}@hpi.de

Abstract The majority of web assets cannot be understood by machines, because of the lack of available explicit and machine readable semantics. By enabling machines to understand the meaning of web media, fully automated discovery, processing, and linking become feasible. Semantic Web technologies offer the possibility to enhance web resources with explicit semantics via linking to ontologies encoded in RDF. We demand to make the content of every web asset explicit for machines with the least possible effort for any content provider. Web servers should deliver RDF descriptions for any web document on request. To achieve this, we propose a framework that enables web content providers to connect to content-wise descriptions of their web assets via simple HTTP content negotiation in connection with on-the-fly automated multimedia analysis services. We demonstrate the feasibility of our approach with a prototype implementation.

Keywords: machine understandability, web media, automated media analysis, semantic web technologies, RDF, content negotiation

1 Introduction

The Web is made for humans, not for machines. The majority of web assets cannot be understood by machines, because of the lack of available explicit and machine readable semantics. To fully automatically discover, process, and link web content, machines must be able to understand its meaning. Nowadays, multimedia documents such as images, video and audio files, but also other electronic documents such as PDFs, various formats for word processors, spreadsheets, slide show presentations, and file archives are indispensable constituents of the Web and use up the majority of the available bandwidth in the Internet. These documents largely contain unstructured data, partly in proprietary formats, which makes it intricate for machines to extract the actual content and meaning. Even though a web browser can display an image, it cannot understand the image content.

Consider the following scenario: someone uploads a holiday photograph to a web server so that it is publicly available to her friends. Those can download the image and admire her in front of that spectacular sight. But, if the image is downloaded by a computer it cannot see or recognize the content of the photograph like a human. Given explicit metadata for that image the computer

would know where the picture has been taken, which objects can be seen, etc. Using this knowledge, a machine could provide background information to the user, link it to the personal data of the user, make it retrievable by its content, and suggest to make use of the image for a particular purpose, e. g. as an illustration in a travel blog.

The Semantic Web [3] introduces languages such as the Resource Description Format (RDF) and the Web Ontology Language (OWL) to bring structure to the content of web pages with the goal to provide explicit and machine understandable semantics. One way to provide explicit semantics in HTML pages is the inclusion of microdata, such as RDFa [16] and schema.org[1], to annotate web documents with formal descriptions which are connected with the help of vocabularies to Linked Data resources.

Web documents are delivered via the Hypertext Transfer Protocol (HTTP). By using HTTP content negotiation different versions of the same web document can be identified and accessed via one unified URI [4]. To access information resources in the Web of Data, for Linked Data resources the same URI is used to access a human readable HTML document as well as a machine understandable RDF version of the same resource [7]. This mechanism should not be restricted to Linked Data resources only. Content providers should provide content-wise descriptions and metadata for every kind of asset on the Web including multimedia data. Moreover, this should be accomplished with minimal effort, i. e. without an overhead to laboriously create supplementary metadata in a manual way.

When requesting a web asset's URL via HTTP, the computer receives a copy of the original resource. In order to provide a machine understandable explicit semantic description of the web asset, HTTP content negotiation should be enabled and on request an RDF description of the content of the web asset can be delivered. This RDF description can be provided manually, from existing metadata, or with the help of automated analysis algorithms. Overall, the possibility to automatically receive machine readable metadata lowers the barrier for machines to understand and correctly interpret web assets.

In this paper we propose a framework based on standardized web protocols to enable the delivery of machine readable content related metadata for arbitrary documents on the web independent of the web document's type, modality, and encoding. To enable a smooth and least effort delivery of metadata we propose to utilize the content negotiation mechanism that enables to identify the original content as well as its metadata via the same URI. We demonstrate the feasibility of our approach with a prototype implementation that combines automated visual analysis as a web service with the content negotiation and metadata delivery mechanism with little effort for any content provider.

The paper is structured as follows: Sect. 2 describes technologies and description formats related to content representation, followed by potential use cases and service provisioning. Sect. 3 provides a detailed description of the prototypical implementation. Sect. 4 summarizes related approaches and Sect. 5 concludes the paper.

[1] https://schema.org/

2 Content Representation and Content Negotiation

HTTP content negotiation is a well established mechanism that is used to deliver different representations of a document from a web server according to the demands and constraints of the user agent (client). By submitting a request to a web server, the client informs the web server what media types it understands including a ranking of preference. The client provides an HTTP accept header that lists acceptable media types, as e. g. `Accept: text/html`. In general, the accept header lists the MIME Types of the media that the client is willing to process [4]. The web server is then able to supply the version of the resource that best fits the user agent's needs.

In the Web of Data this mechanism is applied to identify resources, i. e. Linked Data resources, with the same URI providing a human readable HTML version as well as a machine readable (or even machine understandable) RDF version [7]. Thereby, various representations of the same information can be delivered using the same URI to identify this information.

2.1 Possible Contents of Descriptions

There exist different types of content descriptions. We classify these types into three different layers: *file metadata*, *provenance information*, and an actual *description of the content* (cf. Table 1). The most generic type of descriptions is *file metadata* which is often already available from the HTTP header. Depending on the file type there might also be *file type specific metadata*, in the case of an image that would be e. g. its pixel dimension and compression rate, for an audio or video file its duration, and for a PDF document the number of pages. A second type which is generically applicable and becomes increasingly relevant is *provenance information* such as the creator, creation and modification dates, and rights information. Images may contain Exif information with technical metadata such as the camera model, shutter speed, and geo-location.

The actual *description of the content* does strongly depend on the file type. For example, images could be described by color-space histograms, image type (e. g. photo, clip-art, line drawing, animated, etc.), or more sophisticated categorization methods, such as visual concept detection [9]. Audio transcriptions extracted from speech recognizers could be shipped along with any audible content. The textual content of such transcriptions as well as from rich text formats could be used to be categorized with extracted keywords or text summaries. Semantic named entity linking [19] could be used to identify meaningful elements in text and provide links to referenced resources.

2.2 Description Formats

For the content-wise description of multimedia documents various metadata schemata and vocabularies have been proposed for which also RDF based versions have been created. The following non-exhaustive list shows some prominent vocabularies within the context of the proposed system:

Table 1. Selection of Multimedia Content Descriptions

Generic	*Image*	*Video*	*Audio*	*PDF (publication)*
File Metadata				
File size	Image Width	Video Width	Codec	Number of Pages
File type	Image Height	Video Height	Duration	Page Size
MIME type	Compression	Codec	Sample Rate	
		Duration		
Provenance Information				
Creator	Camera Model	Camera Model	Artist / Speaker	Author
Rights	Exposure Time		Recorder Model	Publisher
Creation Date	Aperture			DOI
Modification Date	GPS Position			
Content Description				
	Image Classification	Audio Transcript	Audio Transcript	Abstract
	Visual Content Detection	Shot Boundaries	Genre	Keywords
	Face Detection	Spatio-temporal Annotations	Title	Citations
	Object Detection		Album	Experimental Data

- **DC Element Set / DC Terms:** The *Dublin Core vocabularies* provide a small set of elements for the description of web resources as well as of physical objects [15].
- **MIME type:** the *Multipurpose Internet Mail Extensions* (MIME) clearly specify multimedia content types as well as content encodings [5].
- **Exif:** *Exchangeable image file format* (Exif) specifies a set of tags to describe image formats and technical metadata of camera and imaging devices. Kanzaki [12] provides an RDF vocabulary to encode Exif picture data [11].
- **COMM:** The *core ontology for multimedia* (COMM) [1] has been built re-engineering the multimedia annotation standard MPEG-7 [10].
- **Open Annotation Ontology:** the *Open Annotation Data Model* specifies an interoperable framework for creating associations between related resources and annotations [17].
- **NIF:** the *NLP Interchange Format* (NIF) is an RDF/OWL-based format that aims to achieve interoperability between natural language processing (NLP) tools, language resources and annotations on different levels [8].
- **Media Fragments:** the *W3C media fragments recommendation* specifies how to construct media fragment URIs and their utilization with the HTTP protocol [18].

2.3 Service Provisioning

Two options for deployment of such a feature are conceivable: *local* or *external* creation of RDF descriptions. If the content provider decides to host also the RDF descriptions, he has full control over the content and can integrate background knowledge, e. g. from media asset management tools. It would be reasonable to integrate this feature in content management systems. Otherwise, this task can also be transferred to a dedicated service provider, who analyses the file and generates the RDF description. Such a provider might be able to deploy more sophisticated content analysis tools to provide consistent descriptions. We demonstrate the latter approach in Sect. 3 since it allows a very simple setup for any content provider.

2.4 Potential Use Cases

There are plenty of application scenarios conceivable that would benefit from rich descriptions of web media.

- **Hypermedia and Accessibility** Formal descriptions of web assets can support the accessibility for end users who are in any form impaired to perceive the original media format, e. g. a screen reader compiles and reads a natural language description of the visual content to a blind user. Such description does not need to be static as e. g. provided by the `alt` tag for images. Instead, intelligent tools could generate textual content-wise descriptions from sophisticated visual analysis results. Moreover, links to related resources can be attached to media files, e. g. sections in an e-learning video could be linked to forum discussions where learners discuss questions raised by the lecturer.
- **Multimodal search, SEO, and Recommender systems** Search engine support within multimedia and other unstructured files is hard to achieve. A search engine needs to analyze the data first in order to index it properly. While big search providers can operate the needed infrastructure, i. e. computing power and algorithms, enterprise search engines are able to provide multimodal search based on the media's content descriptions. But also web scale search engines might provide better search results by using this explicit information and honor the provision of such.
- **Generic API** Instead of developing and deploying new APIs for the distribution of metadata or content descriptions, content negotiation and RDF can be used to deliver such information in a generic way. E. g. video transcripts or subtitles are currently provided as an extra file, via dedicated APIs, or embedded in the video stream itself. Similar holds for chapter marks and shownotes[2] in podcasts. Whatever additional information shall be provided for web assets in future, the suggested mechanism can easily be applied to it.

3 Implementation and Demo

We have set up a prototype implementation that enables the creation and delivery of content-related RDF descriptions for images including basic technical metadata, Exif data, as well as descriptive metadata from automated visual concept detection. The overall architecture principle is depicted in Fig. 1. The demo consists of a standard web server and the content analysis server (COAL). The standard web server is considered to be an ordinary web server hosting some arbitrary website. The purpose of the COAL server is to provide RDF content descriptions for images as a service. The website publisher simply configures the web server to redirect specific content type requests to the COAL server. An exemplary rewrite rule, which redirects RDF data requests for image URLs to an external server and adds an alternate link header, is given in Listing 1.

[2] e. g. as provided at http://shownot.es/

In Fig. 1, the client, e. g. a web browser plugin or a search engine, requests a resource's machine readable RDF description from the web server by specifying the HTTP header field `Accept: application/rdf+xml` (1). The web server applies the rewrite rule and sends an HTTP `303 redirect` back to the client including the new redirect location (URL) pointing to the COAL server (2). The client then requests the given URL from the COAL server (3), which subsequently retrieves the original file from the web server (4, 5) and ingests it to the analysis workflow. The analysis results are encoded as RDF and sent back to the client (7). A standard HTTP cache serves as temporary storage to ensure a resource is analyzed only once within a certain range of time.

```
<FilesMatch "\.(gif|jpg|jpeg|png|GIF|JPG|JPEG|PNG)">
  <IfModule mod_rewrite.c>
    RewriteEngine on
    RewriteRule ^ - [E=ORIGINAL_URI:http://%{HTTP_HOST}%{REQUEST_URI}]
    RewriteCond %{REQUEST_FILENAME} -f
    RewriteCond %{HTTP_ACCEPT} ^.*text/turtle.* [OR]
    RewriteCond %{HTTP_ACCEPT} ^.*application/n-triples.* [OR]
    RewriteCond %{HTTP_ACCEPT} ^.*application/rdf\+xml.* [OR]
    RewriteCond %{HTTP_ACCEPT} ^.*application/ld\+json.*
    RewriteRule . http://coal.s16a.org/resource?url=%{ENV:ORIGINAL_URI} [R
     =303,L]
    <IfModule mod_headers.c>
      Header append Link "<http://coal.s16a.org/resource?url=%{ORIGINAL_URI}e
     >; rel=\"alternate\"; type=\"application/rdf+xml\""
    </IfModule>
  </IfModule>
</FilesMatch>
```

Listing 1. Apache rewrite rule to redirect to the COAL server

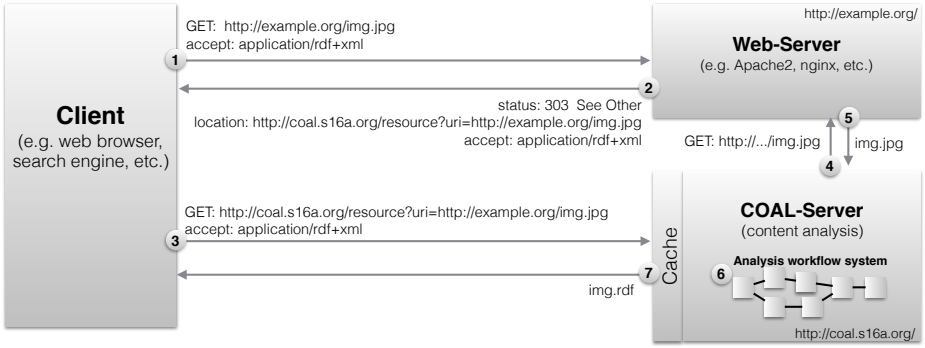

Figure 1. Principle of content analysis with content negotiation for a given image

We have configured the rewrite rule on our Wordpress-based blog[3]. Image content descriptions can now easily be requested by specifying the desired content type:

```
curl -L "http://blog.yovisto.com/wp-content/uploads/2015/07/Bumper8.jpg" -
H "Accept:application/rdf+xml"
```

[3] http://blog.yovisto.com/

4 Related Approaches

The initial idea seemed so obvious that we did not expect not to find anyone who had at least tried it before, and indeed: already in 2002, Lafon and Bos released a W3C note [13] for describing photos with RDF and HTTP content negotiation. The approach of *Photo RDF* includes a manual annotation of digital images, supported by the *rdfpic* data entry program, and an extension for the Jigsaw server. A demonstration server for *Photo RDF* is also available[4]. *Photo RDF* already comes close to our vision, but it is limited to images and relies on manual annotation, which might be the main reason for its limited use.

The Adobe Extensible Metadata Platform (XMP)[5] allows to embed RDF descriptions in the file header of several file formats. Adobe recommends to use the Dublin Core vocabulary for provenance information and offers additional schemas. XMP is supported by a number of tools.

Semantic annotation of multimedia has been a field of research for over a decade. The goal is to provide rich machine processable descriptions of media contents using well defined properties. A number of models and tools have been created [14]. The most commonly used vocabularies are the W3C Media Ontology [2] and the Open Annotation Model [17]. Temporal and/or spatial regions in media assets are referenced via Media Fragment URIs [18]. These activities usually focus on individual collections and have not been applied at web scale. Furthermore, there is no common mode of publication for such media annotations, while content negotiation has been suggested [6], others such as SPARQL endpoints or individual APIs are also used.

DBpedia Commons[6] provides RDF descriptions for Wikimedia Commons including its multimedia resources [20]. The descriptions are extracted from the Wikimedia Commons wiki pages using the DBpedia extraction framework, i. e. they mainly include handcrafted annotations while low-level file information is not contained. Unfortunately, the data is not linked by its original source via standard HTTP protocols.

5 Conclusion and Outlook on Future Work

In this paper we have sketched our vision to realize a machine understandable web of media assets, which bases entirely on state-of-the-art web technologies and to a great extent can be implemented in an automated way. We provided a number of use-cases that would benefit from explicit media content descriptions or are becoming possible by that. As always, it demands a significant amount of deployments to get real use of it. We have demonstrated that the actual deployment can be as easy as pie by using dedicated services.

Still, there are steps to take: a common set of ontologies to describe web assets and their content needs to be agreed. Furthermore, we plan to extend the

[4] http://jigsaw.w3.org/Yves/Australia/1998/04/
[5] http://www.adobe.com/products/xmp.html
[6] http://commons.dbpedia.org/

COAL demo implementation in a modular way to support additional media and file types with more sophisticated analysis technologies.

References

1. Arndt, R., Troncy, R., Staab, S., Hardman, L., Vacura, M.: COMM: Designing a well-founded multimedia ontology for the web. In: The Semantic Web – ISWC/ASWC 2007. pp. 30–43 (2007)
2. Bailer, W., et al.: Ontology for media resources 1.0. W3c recommendation, W3C (Feb 2012), https://www.w3.org/TR/mediaont-10/
3. Berners-Lee, T., Hendler, J., Lassila, O.: The semantic web. Scientific American 284(5), 34–43 (2001)
4. Fielding, R.T., et al.: Hypertext transfer protocol – http/1.1. RFC 2616, RFC Editor (June 1999), http://www.rfc-editor.org/rfc/rfc2616.txt
5. Freed, N., Borenstein, D.N.S.: Multipurpose Internet Mail Extensions (MIME) Part One: Format of Internet Message Bodies. IETF RFC 2045 (Mar 2013), https://rfc-editor.org/rfc/rfc2045.txt
6. Hausenblas, M., et al.: Interlinking multimedia. In: Proc. of the Linked Data on the Web Workshop. CEUR-WS, vol. 538. Madrid, Spain (April 2009)
7. Heath, T., Bizer, C.: Linked Data: Evolving the Web into a Global Data Space. Morgan & Claypool, 1st edn. (2011)
8. Hellmann, S., Lehmann, J., Auer, S., Brümmer, M.: Integrating nlp using linked data. In: The Semantic Web – ISWC 2013 (2013)
9. Hentschel, C., Sack, H.: What image classifiers really see – visualizing bag-of-visual words models. In: MultiMedia Modeling: 21st Int. Conf. pp. 95–104. Springer (2015)
10. MPEG-7: Multimedia content description interface (2001)
11. Exchangeable image file format for digital still cameras: Exif version 2.3 (2010), http://home.jeita.or.jp/tsc/std-pdf/CP3451C.pdf
12. Kanzaki, M.: Exif data description vocabulary (2003), http://www.kanzaki.com/ns/exif, last update in 2007
13. Lafon, Y., Bos, B.: Describing and retrieving photos using RDF and HTTP. W3c note, W3C (Apr 2002), https://www.w3.org/TR/photo-rdf/
14. Nixon, L., Troncy, R.: Survey of semantic media annotation tools for the web. In: ESWC 2014 Satellite Events, pp. 100–114. Springer (2014)
15. Powell, A., Nilsson, M., Naeve, A., Johnston, P.: Dublin core metadata initiative - abstract model (2005), http://dublincore.org/documents/abstract-model
16. RDFa 1.1 primer: Rich structured data markup for web documents. Working group note, W3C (03 2015), http://www.w3.org/TR/rdfa-primer/
17. Sanderson, R., Ciccarese, P., de Sompel, H.V.: Open annotation data model. W3c community draft, W3C (Feb 2013), http://www.openannotation.org/spec/core/
18. Troncy, R., et al.: Media fragment URI 1.0. W3c recommendation, W3C (Sep 2012), https://www.w3.org/TR/media-frags/
19. Usbeck, R., et al.: GERBIL – general entity annotation benchmark framework. In: 24th WWW conference (2015)
20. Vaidya, G., Kontokostas, D., Knuth, M., Lehmann, J., Hellmann, S.: DBpedia Commons: Structured multimedia metadata from the wikimedia commons. In: Arenas, M. (ed.) The Semantic Web – ISWC 2015. pp. 281–289. Springer (2015)