V4Ann: Representation and Interlinking of Atom-based Annotations of Digital Content

Georgios Meditskos, Stefanos Vrochidis, Ioannis Kompatsiaris
Outline

• Overview & Motivation
• Proposed Framework
  • V4Ann Annotation Model
  • Inference and Validation
  • Asset searching
• Evaluation
• Future Directions
Overview & Motivation

• **High quality content** is nowadays widely available on the web and other sources
  • user-generated content, such as images, videos and text posted by users on social media, wikis and blogs
  • content provided through official publishers and distributors, such as digital libraries, organisations and online museums

• This content remains largely **under-exploited**
  • lack of solutions for its retrieval and integration into the design process

• If leveraged appropriately, could serve as a **valuable source of inspiration**
  • a great source of revenue for the Creative Industries, such as architecture and video game design
  • inspire and support the creation of new content and to produce new assets or to enhance and (re-)use the already existing ones
Challenge

• Maximise the potential for re-purposing of digital content
• Development of innovative technologies to systematically
  • Analyse
  • Combine
  • Link
  • Foster searchability and reusability of heterogeneous content
V4Design Project

http://www.v4design.eu/

Visual and textual content re-purposing FOR(4) architecture, Design and video virtual reality games
V4Design Concept

• Vision
  • Reuse and repurpose high quality content
  • Provide repurposed content to targeted creative industries
    • Architects, VR and video game designers
  • Provide revenues for the data providers and creative industries

• Key Technologies
  • Web data mining for crawling visual and textual data
  • Aesthetics extraction and texture proposals
  • Language understanding and text generation
  • 3D reconstruction
  • Semantic knowledge representation, linking and reasoning
Overview of V4Design Platform
V4Ann Role in V4Design

• Semantic middleware, capturing, interlinking and serving analysis results to multimedia analysis services
  • Annotation propagation and linking: efficient and interoperable way to represent, exchange and further link metadata, both structurally and semantically

• Semantic atom-based query infrastructure to retrieve generated assets
  • Context-aware retrieval: practical and efficient retrieval mechanisms on top of the multimodal annotations

• How?
  • Web Annotation Data Model (WADM), Building Topology Ontology (BOT)
  • Domain-specific ontologies (Europeana Data Model)
  • SPIN / SHACL inference and validation rules
V4Ann Annotation Model
Key Concepts

- Annotations
  - resource containers, implementing the annotation pattern of WADM

- Media types
  - image, video, text, 3D model

- Views
  - container for attaching annotations (atoms) to media types

- Atoms
  - aesthetics, object/building types, named entities/relations, image/video frames used for 3D reconstruction
Annotation resources

- LocalisationAnnotation, TextualAnnotation, AestheticsAnnotation and 3DModelAnnotation
- Extends oa:Annotation concept
  - hasContext ⊑ oa:hasBody
  - describes ⊑ oa:hasTarget
- V4Ann annotation has a context that describes a media type using views

Annotation ⊑ oa:Annotation ⊑
∃describes.MediaType ⊑ ∀hasContext.View
Media types

• Four media types
  • Video
  • Text
  • Image
    • Mask ∈ Image
    • Texture ∈ Image
  • 3DModel

• Intuitively, each media type resource represents a single multimedia asset for which a set of annotation atoms needs to be captured.
Views and Atoms

• Encapsulate the annotation data (for media types)
  • Annotation are derived from content analysis (text analysis, visual analysis, etc.)

• **Aesthetics**: categorisation of the aesthetics of paintings and images that contain architecture objects and buildings
  • Style: e.g. impressionism, cubism and expressionism
  • Creator: mainly for paintings (schema.org)
  • Emotion: e.g. fear

\[
\text{AestheticsAnnotation} \sqsubseteq \text{oa:Annotation} \sqcap \\
\exists \text{describes.\{} \text{Image} \sqcap \text{Video} \text{\}} \sqcap \forall \text{hasContext.AestheticsView} \\
\text{AestheticsView} \sqsubseteq \forall \text{creator.Creator} \sqcap \forall \text{style.Style}
\]
Views and Atoms

• Object and Building Localisation: Building and interior objects localization, e.g. include tables, vases, as well as statues, buildings, etc.
  • Masks: outline of the object
  • Tags: computer vision labelling
  • Frames

\[
\text{LocalisationAnnotation} \subseteq \text{oa:Annotation} \sqsubseteq \\
\exists \text{describes.\{Image, Video\}} \sqsubseteq \forall \text{hasContext.LocalisationView} \\
\text{LocalisationView} \subseteq \exists \text{hasTag.Tag} \sqsubseteq \forall \text{hasFrame.integer}
\]
Views and Atoms

• **Text Analysis**: annotation enrichment with entities and concepts extracted from titles, captions, descriptions, etc.
  • Already disambiguated: WordNet, BabelNet or Dbpedia

```
TextAnnotation ⊑ oa:Annotation ⊓
∃describes.{Image ⊔ Video} ⊓ ∀hasContext.TextView

TextView ⊑ ∃hasTag.Tag
```
Views and Atoms

• **3D reconstruction**: Converts input videos / images into 3D point clouds and meshes
  - 3D-related properties (e.g. number of points )
  - Source of 3D reconstruction (very important for annotation propagation and linking!)

\[ \text{3DModelAnnotation} \sqsubseteq \text{oa:Annotation} \sqcap \]
\[ \exists \text{describes.3DModel} \sqcap \forall \text{hasContext.3DModelView} \]
\[ \text{3DModelView} \sqsubseteq \exists \text{hasSource.\{Images \sqcup Video\}} \]
Inference and Validation
Implicit Relations

• Additional inferences are derived by combining native OWL 2 RL reasoning and custom rules
  • SPARQL-based CONSTRUCT graph patterns are used that identify the valid inferences that can be made on the annotation graphs.

• Example: atom propagation
  • Propagate / interlink atoms among view-based annotation provided that they are somehow connected
  • For example, the aesthetics atoms extracted from video frames can be used to annotate the 3D models that have been reconstructed using those frames

```sparql
CONSTRUCT {  
} WHERE {  
  ?a1 a :AestheticsAnnotation;  
  ?a2 a :3DModelAnnotation; :hasContext ?view .  
}
```
Validation and Consistency Checking

• Consistency, structural and syntactic quality of the metadata

A. Native ontology consistency checking (e.g. OWL 2 DL reasoning)
   • TBox consistency (e.g. class disjointness)

B. Custom SHACL validation rules
   • constraint violations, e.g. missing values and cardinality violations
     • e.g. that all 3D model views should include references to the atoms (images) used for the 3D reconstruction.

```sparql
v4d:3DModelView
  rdf:type sh:NodeShape ;
  sh:property [ 
    rdf:type sh:PropertyShape ; sh:path v4d:image ;
    sh:class v4d:MediaType ; sh:minCount 1 ;
    sh:name "one or more images" ; sh:nodeKind sh:IRI ; 
  ] .
```
Context-based Asset Retrieval
Local Context

• Enriched, pre-constructed semantic signature of this atom
  • conceptual and lexical relations from existing semantic networks and datasets, such as WordNet, BabelNet and ConceptNet

• The retrieval mechanism aims to match incoming local contexts of query atoms (keywords) against local contexts of annotation atoms
Atom Similarity (AH Metric)

• Similarity of two atoms taking into account their local context
  • Term similarity function $S(A, B) \in [0, 1]$
  • Set $F$ of local context filters

• Filters ( $A \overset{f}{\sim} B$ )
  • exact: The two atoms should have either the same URI, or they should be equivalent concepts
  • plugin: The atom B should belong to the set of hypernyms of A or to the set of relevant concepts of A
  • subsume: The atom B should belong to the set of the hyponyms of A

• The atom A matches the atom B, with respect to a filter set $F$, if and only if there is at least one filter $f \in F$

$$A \overset{F}{\sim} B \iff \exists f \in F : A \overset{f}{\sim} B$$
S Function

- Heuristic function that takes into account the information capture in local contexts

\[ r_1: \text{if } A = B \lor A \equiv B, \text{ then } S(A, B) = 1. \]
\[ r_2: \text{if } B \in h y_A \lor B \in r_A, \text{ then } S(A, B) = a. \]
\[ r_3: \text{if } B \in h o_A, \text{ then } S(A, B) = b. \]
\[ r_4: S(A, B) = 0. \]

- a and b (a > b) are defined manually based on domain knowledge regarding the quality of multimedia analysis that produces the atoms (e.g. aesthetics extraction)

- The empirical dentition of these values (currently a = 0.7 and b = 0.3) aims to promote plugin matches (r2) over subsumed (r3).
Evaluation
Digital Content

- DeutscheWelle (DW) and Europeana are two key content providers
  - DW provides their documentary and movie archives.
  - Europeana provides their large archive of paintings, pictures of contemporary artwork and related critics.
- The generated V4Ann annotation graphs contain the atoms that have been extracted from the analysis components, along with interconnections among the annotation resources.

<table>
<thead>
<tr>
<th>#annotations</th>
<th>#atoms</th>
<th>avg. local context size</th>
</tr>
</thead>
<tbody>
<tr>
<td>17245</td>
<td>154610</td>
<td>17 per atom</td>
</tr>
</tbody>
</table>

"description": "Our drone shows you the Bauhaus University in Weimar, Thuringia."
User-centred Evaluation

1. Collect qualitative feedback on the results, as well as on non-functional aspects, such as query response time

2. (and most important) to generate an annotation dataset and assess the performance of V4Ann

- A list of relevant resources has been provided, such as square names, monuments, building types, etc., in order to help them conduct relevant queries.

- Users filled in a five-point scale questionnaire (1-completely agree, 5-completely disagree).
User-centred Evaluation

• Quality of atoms: The quality and relevance of local contexts depends on the performance of content analysis, e.g. visual and textual analysis. Visual analysis provides, in principle, better results than text analysis.

• Retrieval results: The system achieves good top-ranked accuracy, however the complete set of the results contain quite a lot irrelevant entries.

• Response time: The response time of the system was positively assessed. The average response time was 4.1 seconds, which includes query analysis, building of local context and search algorithm execution.
System Evaluation

• As expected, the stricter the filter is, the more accurate results we obtain (high precision) with low, however, recall

• A higher h value leads to more generic local contexts that affect precision
  • For example, the third-level WordNet hypernym of “tower” is “unit”, which is too generic

• The more specific the label/atom is, the more room for additional context exists

<table>
<thead>
<tr>
<th></th>
<th>$h = 1$</th>
<th></th>
<th>$h = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Recall</td>
<td>Precision</td>
<td>Recall</td>
</tr>
<tr>
<td>exact</td>
<td>0.59</td>
<td>0.77</td>
<td>0.44</td>
</tr>
<tr>
<td>plugin</td>
<td>0.67</td>
<td>0.69</td>
<td>0.52</td>
</tr>
<tr>
<td>subsume</td>
<td><strong>0.73</strong></td>
<td>0.61</td>
<td>0.59</td>
</tr>
</tbody>
</table>
Conclusion

• Ontology-based framework for representing, linking and enriching results of multimedia analysis on digital content

• Reuses existing standards and schemata, building the atom-based annotations graphs on top of standard ontologies, controlled vocabularies and patterns
  • WADM pattern

• We evaluated the framework using actual multimedia content and atoms provided by the V4Design modules

• V4Ann is accessible through Rhinoceros 3D and Unity plugins developed in the V4Design project.
Future Directions

• Implement context-aware algorithms to improve the classification accuracy of incoming atoms
  • For example, if the wrong style for a painting is provided by aesthetics, this will affect precision, since V4Ann does not aim at improving the classification of incoming atoms

• Extend the context-aware retrieval algorithm with more sophisticated similarity metrics and functions
Thank you!

gmeditsk@iti.gr
https://mklab.iti.gr

http://www.v4design.eu/