

IMGpedia: A Proposal to Enrich DBpedia with Image Meta-Data

Benjamin Bustos and Aidan Hogan

Center for Semantic Web Research
Department of Computer Science
University of Chile
{bebustos, ahogan}@dcc.uchile.cl

Abstract. We introduce IMG_{PEDIA}: a research proposal aiming to bridge structured knowledge-bases and multimedia content. Our concrete plan is to enrich DB_{PEDIA} data with further metadata about images from WIKIPEDIA, including content-based visual descriptors. Our concrete goal is to create a unified querying and inference system that allows for interrogating the DB_{PEDIA} knowledge-base and the visual content of WIKIPEDIA's images together. Our broader ambition is to explore methods by which multimedia data can be made a first-class citizen of the Semantic Web.

1 Introduction

DB_{PEDIA} [1] is an ongoing effort by the Linked Data community to extract structured content from WIKIPEDIA and represent it in RDF. The main goal is to enable users to query the content of WIKIPEDIA as a whole, getting direct answers automatically aggregated from multiple articles. The most recent version of DB_{PEDIA} contains billions of facts extracted from 125 language versions of WIKIPEDIA, with links to and from dozens of external datasets. Over the past seven years, it has become *the* central dataset of the Linked Open Data community [5].

However, DB_{PEDIA} mainly focuses on extracting information from WIKIPEDIA's info-boxes: attribute–value panes that appear on the right-hand side of articles. As such, aside from adding links, DB_{PEDIA} ignores the images appearing in the body of the article for a given entity as well as the structured data available in image pages: no meta-data are extracted for images. Like many initiatives in the Semantic Web¹, DB_{PEDIA} links to but otherwise disregards multimedia content.

Our proposal is thus to extract and associate meta-data from the images embedded in WIKIPEDIA and link the resulting corpus with the DB_{PEDIA} dataset. This dataset – which we call IMG_{PEDIA} – would consider all images in an article, all meta-data associated with the image available from WIKIMEDIA (author, date, size, etc.) and visual descriptors that capture the content of the image itself.

¹ But not all: see, e.g., <http://www.w3.org/2005/Incubator/mmsem/>

We are motivated by the idea of creating a corpus that allows for querying, in unison, both the structured/semantic meta-data of DBPEDIA and the visual content extracted from images; e.g., “*give me Europe cathedrals that have an image visually similar to one of the external images for Cusco Cathedral in Peru*”. Likewise, we foresee the possibility of inferring new links from this dataset, e.g., inferring that Saddam Hussein and Donald Rumsfeld have met based on being associated with the same image (in which they are co-depicted). The resulting corpus may also serve as an interesting experimental dataset for the image-processing community, where the structured data associated with images may serve as a ground truth.

2 Images and Visual Descriptors

Before describing IMGPEIDIA, we need to introduce some basic concepts about how images are encoded and what are visual descriptors. An image is a matrix of so-called *pixels* (picture elements). A pixel contains information about its color, which can be displayed for example on a computer monitor. There are several ways to encode the color information of a pixel, which depends on the selection of a color space. Common color spaces are RGB (red-green-blue, used by computer monitors) and CMYK (cyan-magenta-blue-black, used by printers), where colors are represented as tuples of numbers; for example, an RGB color is represented by a three tuple. There are several ways to compress the image encoding, mainly lossy compression methods (e.g., JPEG format) and lossless methods (e.g., PNG format).

A visual descriptor is a way of characterizing an image based on its content. This can be done considering the whole image (global descriptor) or regions of interest detected on the image (local descriptors). For this work, we will initially focus on global descriptors since they can be computed more efficiently than local descriptors, and likewise similarity between them is also more efficient to compute.

Visual descriptors can be defined in several ways; e.g., based on the colours, texture and/or shape of the image. They do not include any semantic information about what appears on the image—hence why they are also called “low-level features”. For instance, a simple colour descriptor is the colour histogram [2], that captures the distribution of colour in the image. We note that visual descriptors are usually vectors of high dimensionality (tens to hundreds of real values).

Visual descriptors allow us to implement, e.g., content-based similarity search. A similarity query in an image data set returns the most similar images, according to its content, to a given one (the query image). This is also known as *query-by-example*. Formally, let \mathbb{U} be the universe of all images, let $\mathbb{S} \in \mathbb{U}$ be the image data set, and let $\delta : \mathbb{U} \times \mathbb{U} \rightarrow \mathbb{R}^+$ be a function (the distance) that returns how dissimilar are two images. There are two basic types of similarity queries: (1) *Range query*: given the query image $q \in \mathbb{U}$ and a tolerance radius $r \in \mathbb{R}^+$, return all images from \mathbb{S} that are within distance r to q ; (2) *Top-k query*: return the k -closest objects to q . If \mathbb{S} is formed by all visual descriptors

(high-dimensional vectors) extracted from the images in the data set, and if q is the visual descriptor of the query image, and if δ is any metric function (e.g., the Euclidean distance), it is relatively straightforward to implement content-based range and top- k queries over \mathbb{S} .

3 IMGpedia Dataset

Our vision of IMGpedia is an enhanced version of DBpedia with image entities. An image entity contains metadata (e.g., title, subject, source, format, description, date, size, location, etc.) and content-based descriptors (e.g., colour descriptor) of the image. Image entities can be linked with other entities (not necessarily images).

For creating the IMGpedia dataset itself, we propose the following procedure:

- Locate and download images/image-pages from WIKIMEDIA.
- Extract meta-data from the image page, including its size, author, licence, etc. Annotate images with tags computed from its (possibly many) captions [4].
- Compute the visual descriptors for the images. For this, we can use global visual descriptors like colour and edge [2], following the MPEG-7 standard [3].
- Create the image entities using the extracted metadata and content-based data.
- Represent and publish the IMGpedia dataset as Linked Data.

4 Querying IMGpedia

Our main research goal is to investigate methods by which semantic data (in this case DBpedia) and multimedia data (in this case describing WIKIPEDIA images) can be combined such that they can be queried in a holistic manner. In the context of IMGpedia, our approach is divided into three main parts: materialising links between image resources, extending SPARQL to execute content-based analysis at runtime, and inferring new links between “primary entities” based on image data.

Materialising relations between images using content-based descriptors. Low-level descriptors do not contain any semantic information about the original image, making them hard for users to leverage in queries. This problem is known as the *semantic gap* [6]. However, high-level relations among image entities can be computed from visual descriptors and similarity queries. For example, the relation **near-copy** can be defined as two different images with distance δ less than some threshold τ . By using range queries, it would be easy to find all pairs of near-copies among the images. Other relevant relations that can be considered are **alt-size**, **contains** and **similar**. These could also be materialised as triples and added to the structured knowledge-base, with appropriate inference – e.g.,

for symmetry, reflexivity or subsumption of relations – allowing users to specify SPARQL queries such as:

```
SELECT ?usPolitician WHERE {
  db:Saddam_Husseini foaf:depiction ?img1 .
  ?usPolitician dbo:party db:Republican_Party_(US) ;
    foaf:depiction ?img2 .
  ?img1 i:nearCopy ?img2 .
}
```

Extend SPARQL with functions for content-based image search. Not all content-based user requirements can be anticipated in the form of discrete relations. Hence we propose to extend SPARQL to include content-based analysis features. More specifically, we propose to use extensible functions in SPARQL and custom datatypes to enable queries that combine querying of semantic content and image content. Taking the introductory example, let's say that the user wishes to find cathedrals in Europe with similar images to external images of Cusco Cathedral in Lima:

```
SELECT ?cathedral ?sim WHERE {
  db:Cusco_Cathedral foaf:depiction ?img1 .
  FILTER(i:colorRatio(?img1,i:rgb(40,100,150),i:rgb(170,200,255)) > 0.2)
  ?eurCathedral rdf:type dbo:ReligiousBuilding ;
    dbo:location [ dcterms:subject dbc:Countries_in_Europe ] ;
    foaf:depiction ?img2 .
  BIND(i:sim(?img1,?img2) as ?sim) FILTER(?sim > 0.7)
} ORDER BY ?sim
```

The first `FILTER` uses extended functions to only consider images that have more than 20% of their pixels falling within the cuboid of colours bounded by the two RGB points (looking for blue sky). The subsequent `BIND` and `FILTER` allow the images from European buildings to be filtered and ordered by similarity.

A major challenge here is balancing expressivity and efficiency. In the above case, given a reasonable query plan, the first filter can be applied over the six images appearing in the Cusco Cathedral article, but then all images of all religious buildings in Europe need to be compared with the images that pass the first step. In order to improve the performance of queries, we propose to investigate the use of image indexing techniques that allow for such filters to be executed a lookup, rather than a post-filter, which should lead to more options for query planning. For example, in the query above, a more efficient query plan may try to bind values for `?img2` using a similarity range query (over values bound for `?img1`) allowing for a join to be computed with the knowledge-base rather than applying a brute-force similarity filter over bindings produced by the knowledge-base for `?img2`.

We see this as being one of the deepest technical challenges posed by the work: creating cost models and query plans that combine indexes over the knowledge-base and multimedia content appears to be a challenging but general problem.

Content-based-driven knowledge discovery. A more speculative idea is to infer new knowledge about the data using the images entities and their relations. For example, say that two DBPEDIA resources are associated with the same (near-copy of an) image. If both resources are of type `dbo:Person`, the relation `hasMet` could be inferred. If one resource was a `dbo:Person` and the other was a `dbo:Place`, the relation `hasVisited` could be inferred. Such inferences could be axiomatised as domain-specific rules. Of course, the resulting inferences may not always be crisp conclusions, but may be associated with a confidence value.

5 Conclusions

In this short paper, we have introduced and motivated IMGPEdia: a proposal to enrich DBPEDIA with meta-data extracted from WIKIPEDIA images. We view IMGPEdia as a concrete use-case through which to investigate the challenges and opportunities of combining semantic knowledge-bases with multimedia content.

Acknowledgements This work was supported by the Millennium Nucleus Center for Semantic Web Research, Grant № NC120004, and Fondecyt, Grant № 11140900.

References

1. J. Lehmann, R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P. N. Mendes, S. Hellmann, M. Morsey, P. van Kleef, S. Auer, and C. Bizer. DBpedia - A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia. *Semantic Web Journal*, 2014.
2. B. S. Manjunath, J.-R. Ohm, V. V. Vasudevan, and A. Yamada. Color and texture descriptors. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(6):703–715, 2001.
3. MPEG-7 Overview. URL: <http://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm> (accessed: 2015-01-29), 2015.
4. S. Noah, D. Ali, A. Alhadi, and J. Kassim. Going Beyond the Surrounding Text to Semantically Annotate and Search Digital Images. In *Intelligent Information and Database Systems*, pages 169–179. 2010.
5. M. Schmachtenberg, C. Bizer, A. Jentzsch, and R. Cyganiak. Linking Open Data Cloud Diagram 2014. <http://lod-cloud.net/>; l.a. 2015/01/30.
6. A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, 2000.