

Querying Wikimedia Images using Wikidata Facts

Sebastián Ferrada

Institute for the Foundations of Data
DCC, Universidad de Chile
Santiago, Chile
sferrada@dcc.uchile.cl

Benjamin Bustos

Institute for the Foundations of Data
DCC, Universidad de Chile
Santiago, Chile
bebustos@dcc.uchile.cl

Nicolás Bravo

DCC, Universidad de Chile
Santiago, Chile
nbravo@dcc.uchile.cl

Aidan Hogan

Institute for the Foundations of Data
DCC, Universidad de Chile
Santiago, Chile
ahogan@dcc.uchile.cl

ABSTRACT

Despite its importance to the Web, multimedia content is often neglected when building and designing knowledge-bases: though descriptive metadata and links are often provided for images, video, etc., the multimedia content itself is often treated as opaque and is rarely analysed. IMGPEDIA is an effort to bring together the images of WIKIMEDIA COMMONS (including visual information), and relevant knowledge-bases such as WIKIDATA and DBPEDIA. The result is a knowledge-base that incorporates similarity relations between the images based on visual descriptors, as well as links to the resources of WIKIDATA and DBPEDIA that relate to the image. Using the IMGPEDIA SPARQL endpoint, it is then possible to perform visuo-semantic queries, combining the semantic facts extracted from the external resources and the similarity relations of the images. This paper presents a new web interface to browse and explore the dataset of IMGPEDIA in a more friendly manner, as well as new visuo-semantic queries that can be answered using 6 million recently added links from IMGPEDIA to WIKIDATA. We also discuss future directions we foresee for the IMGPEDIA project.

CCS CONCEPTS

• Information systems → Multimedia databases; Wikis;

KEYWORDS

Multimedia, Linked Data, Wikimedia, Wikidata, IMGPedia

ACM Reference Format:

Sebastián Ferrada, Nicolás Bravo, Benjamin Bustos, and Aidan Hogan. 2018. Querying Wikimedia Images using Wikidata Facts. In *WWW '18 Companion: The 2018 Web Conference Companion, April 23–27, 2018, Lyon, France*. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/3184558.3191646>

1 INTRODUCTION

Multimedia Retrieval and the Web of Data have largely remained separate areas of research with little work in their intersection.

This paper is published under the Creative Commons Attribution 4.0 International (CC BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.

WWW '18 Companion, April 23–27, 2018, Lyon, France

© 2018 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC BY 4.0 License.

ACM ISBN 978-1-4503-5640-4/18/04.

<https://doi.org/10.1145/3184558.3191646>

However, the benefits of working in the intersection of these two areas are clear: the goal of Multimedia Retrieval is to find videos, images, etc., based on analyses of their context and content, while the Web of Data aims to structure the content of the Web to better automate various complex tasks (including retrieval). Further observing that multimedia content is forming an ever-more visible component of the Web, we argue that work in this intersection could prove very fruitful, both to help bring Web of Data techniques more into the mainstream enabling more tangible applications; as well as allowing more advanced Multimedia Retrieval using the knowledge-bases, query languages and reasoning techniques that the Web of Data already provides. Such a combination would allow, for example, to ask semantic queries on the visual content and real-world context of the Web's images, linking media to related sources of information, events, news articles, or indeed other media.

Along these lines, in previous work we initially proposed the IMGPEDIA [4] knowledge-base: a linked dataset that provides information about 15 million images from the WIKIMEDIA COMMONS collection. The dataset of IMGPEDIA includes different visual descriptors that capture various features from the images, such as colour distribution, shape patterns and greyscale intensities. It also provides static similarity relations among the images and links to related entities on DBPEDIA [6] and (now) to WIKIDATA [11]. With IMGPEDIA it is then possible to answer *visuo-semantic* queries through a public SPARQL endpoint. These queries combine the similarity relations with the semantic facts that can be extracted from the linked sources; an example of such a query might be “*retrieve images of museums that look similar to European cathedrals.*”

In our previous work [4] we described the creation of the dataset of IMGPEDIA and illustrated some preliminary queries that it can respond to using the links provided to DBPEDIA. In this paper we report on some new visuo-semantic queries that are enabled by newly added links to WIKIDATA, which provides a more flexible way to request for external entities. We also present a new user interface for IMGPEDIA that allows people to browse the results of the SPARQL queries featuring the images involved, and to explore details about the images, such as their related web resources and their similar images. Finally we present some future directions in which we foresee the IMGPEDIA project developing and possible ways in which it might contribute to existing WIKIMEDIA projects.

2 RELATED WORK

There have been a number of works in the intersection of the Semantic Web and Multimedia areas, and like IMGPEDIA, there are some knowledge-bases based on Semantic Web standards that incorporate – or even focus on – multimedia content. DBPEDIA COMMONS automatically extracts metadata from the media of WIKIMEDIA COMMONS providing triples for licensing, file extension and annotations [10]. Bechhofer et al. [2] provide a linked dataset of live music archives, containing feature analysis and links to existing musical and geographical resources. The MIDI dataset [7] represents multiple music collections in the form of Linked Data using MIDI versions of songs curated by the authors or provided by the community. There are also manually curated datasets such as LINKEDMDB providing facts about movies, and BBC MUSIC [8] describing bands, records and songs. IMGPEDIA [4] is a related effort along these lines but with a current focus on describing images; however, unlike (e.g.) DBPEDIA COMMONS, which focuses purely on contextual meta-data, IMGPEDIA aims to leverage the visual content of images to create new semantic links in the knowledge-base that go beyond surface meta-data.

3 IMGPEDIA OVERVIEW

Before we present novel aspects of IMGPEDIA – the user interface and the new queries enabled by links to WIKIDATA – we first provide an overview of IMGPEDIA, the images from which it is built, the visual descriptors used, the types of relations considered, and so forth. Here our goal is to provide an overview of the knowledge-base; for more details we refer the reader to our previous paper [4].

IMGPEDIA contains information about 14.7 million images taken from the WIKIMEDIA COMMONS multimedia dataset. We only consider the images with extensions JPG and PNG (92% of the dataset) in order to perform a standardised analysis of them. We store the images locally and characterise them by extracting their visual descriptors, which are high-dimensional vectors that capture different features of the images. The descriptors used were the Histogram of Oriented Gradients (HOG), the Gray Histogram Descriptor (GHD), and the Color Layout Descriptor (CLD). These descriptors extract features related to the borders of the image, to the brightness of the greyscale image, and to the distribution of the colour in the image, respectively. Using these sets of vectors we computed static similarity relations among the images: for each image and descriptor, we find its 10 nearest neighbours and we store these relations along with the computed distances and the type of descriptor being used (to scale to the full image-set, we use an approximate nearest neighbour algorithm). Finally, we take all this information and represent it as an RDF Graph, upload it to a Virtuoso Server and provide a public SPARQL endpoint for clients to issue queries. In Table 1 we present some statistics to give an idea of the size of the dataset and to show the main entities and properties provided by IMGPEDIA.

4 BROWSING THE DATA

The initial release of IMGPEDIA used the default interfaces from Virtuoso Server¹ for writing queries, browsing the results and exploring the resources. Though these interfaces provided the necessary mechanisms for agents to access the resources – a SPARQL

endpoint with Linked Data dereferencing – the interfaces for human users were mainly plain HTML tables for displaying SPARQL query results and almost illegible HTML documents for displaying the details of individual resources. We foresaw that the lack of more human-friendly interfaces would prevent people from using and querying the dataset; conversely, being a knowledge-base centred around images, we foresaw much potential for creating a visually-appealing querying and browsing interface over IMGPEDIA.

Along these lines, to initially address the usability of IMGPEDIA, we set up a front-end application that makes the process of querying and browsing our data a more friendly experience. The application was developed with AngularJS 5 framework and uses the Virtuoso Server as a back-end.² The application consists of three main components: a SPARQL query editor, a browser for the results of SPARQL queries that shows the images in-place, and an interface for exploring the details of individual visual entities.

The SPARQL query editor³ provides a text area for writing the queries and a button for executing them. The interface then displays the results of the SPARQL queries below the query editor; if the result contains any visual entities, rather than simply display the URL of the image, it automatically displays the corresponding image by making a request to WIKIMEDIA COMMONS. In Figure 1 we present a screenshot of the interface showing a simple query for three pairs of similar images within a particular visual similarity distance, with the results displayed below.

A user may click on an image displayed in such a result, which will lead them to a detailed focus view of the information available for that visual entity. This view shows the focus image that is being described, along with its name in WIKIMEDIA, and links to related WIKIPEDIA, WIKIDATA and DBPEDIA resources (based on links in those knowledge-bases, appearances in the Wikipedia article associated to those entities, etc.). We also display images similar to the focus image based on precomputed similarity relations present in the IMGPEDIA knowledge-base; similar images are grouped by visual descriptor and are sorted by distance; the user can hover over each such image to see the distance or can click to go to its focus page. In Figure 2 a screenshot of the interface can be found showing a drawing of Queen Mary I of England by Lucas Horenbout; below this the interface displays similar images found through the grey intensity descriptor in ascending order of distance.

5 QUERYING THE DATA

In the first version of IMGPEDIA, the supported visuo-semantic queries relied heavily on the existence of categories for the resources of DBPEDIA; for the example query in Section 1, we use the DBPEDIA category `dbc:Roman_Catholic_cathedrals_in_Europe` for obtaining European Cathedrals (see [4, Figure 3]). To address this issue and in order to support more diverse queries, we have added links to other complementary context sources; in particular, IMGPEDIA now provides 6,614,244 links to WIKIDATA, where a visual entity in IMGPEDIA is linked to an entity from WIKIDATA if the image appears on the English WIKIPEDIA article corresponding to

¹<http://imgpedia.dcc.uchile.cl/sparql>

²The source code of the interface application can be found at <https://github.com/NicolasBravo94/imgpediavis>.

³<http://imgpedia.dcc.uchile.cl/query>

Table 1: High-level statistics for IMGpEDIA

Name	Count	Description
imo:Image	14,765,300	Resources describing individual images
imo:Descriptor	44,295,900	Visual descriptors of the images
imo:similar	473,680,270	Similarity relations
imo:associatedWith	19,297,667	Links to DBpEDIA and WIKIDATA
Total Triples	3,138,266,288	RDF triples present in the IMGpEDIA graph

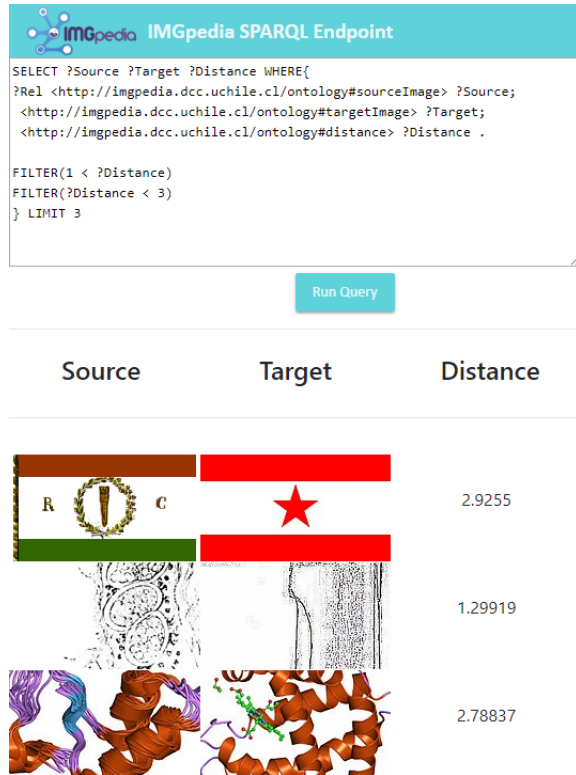


Figure 1: IMGpEDIA SPARQL query editor and results

that WIKIDATA entity. The match between WIKIPEDIA articles and WIKIDATA entities was made querying a dump of WIKIDATA.⁴

Using these new links, we can ask novel queries that were not possible before. For example, now we can request images of governmental palaces in Latin America. This was not possible before since there is no category referring to such palaces in DBpEDIA; if we now leverage these new links to WIKIDATA, we can combine different predicates in order to achieve our goal. In Listing 1 we show the SPARQL query that satisfies our requirements; it first requests from WIKIDATA all entities referring to governmental palaces in Latin America through SPARQL federation, further retrieving the URLs of related images, along with the label of the palace and the name of the country. Figure 3 then shows a sample of the results returned.⁵

⁴The WIKIDATA dump used was that from 2017-07-25

⁵In the figure we can see that there is an image misplaced: the image in the upper right depicts Kedleston Hall in Derbyshire, however in the WIKIPEDIA dump that was

The previous query is what we would refer to as a purely “semantic query”, meaning that it does not rely on any of the visual information present in IMGpEDIA computed from the content of the images. On the other hand, with links to WIKIDATA, we can perform new *visuo-semantic queries* combining WIKIDATA facts and IMGpEDIA visual similarity; for example, we can ask for educational institutions that are similar to the images of government palaces obtained before. In Listing 2 we show the SPARQL query required to do so. First we use the same service clause requesting for the palaces as in Listing 1 so we omit its body here. Later we obtain the similar images and the entities they are associated with and finally we keep those images that are related with educational institutions using SPARQL property paths to consider any subclasses. In Figure 4, we show the result of the query.

In Listing 3 we show another example of a visuo-semantic query where we look for people with images similar to those associated with paintings in the Louvre. In Figure 5 we show the result of the query, where each painting has two similar people. It is worth noting that the first painting is not on display at the Louvre; however the image is related to the Louvre since it is a self-portrait of Jean Auguste Ingres, the painter of *Napoleon I on His Imperial Throne* which is on display at the Louvre, where the portrait of Ingres then appears in the WIKIPEDIA article of the painting.

The SPARQL query of Listing 3 can be modified to obtain other paintings similar to those exhibited at the Louvre by changing the type requested on the second SERVICE clause from wd:Q5 (human) to, for example, wd:Q3305213 (painting). An interesting result of this query variation can be seen in Figure 6, where the two images depict the same painting. In such cases we can say that the images are *near_copy* of each other. Such relations among images could allow reasoning over the entities related to the images; in this case one image corresponds to the painting entity and the other is associated to the painter, and hence it is probable that the painter is the author of the artwork. Section 6 discusses this topic further.

6 FUTURE DIRECTIONS

We presented new features of IMGpEDIA: a more friendly user interface that helps people explore the knowledge-base, and links to WIKIDATA resources that enable novel types of visuo-semantic queries. However, our future goal is to use IMGpEDIA as a starting point to explore and potentially prove a more general concept: that the areas of Multimedia Retrieval and the Web of Data can potentially complement each other in many relevant aspects. Thus

used (2015-12-01) it was stated that said image was used in the English version of the article about the Casa Rosada, and hence it is presented and mislabelled by IMGpEDIA.

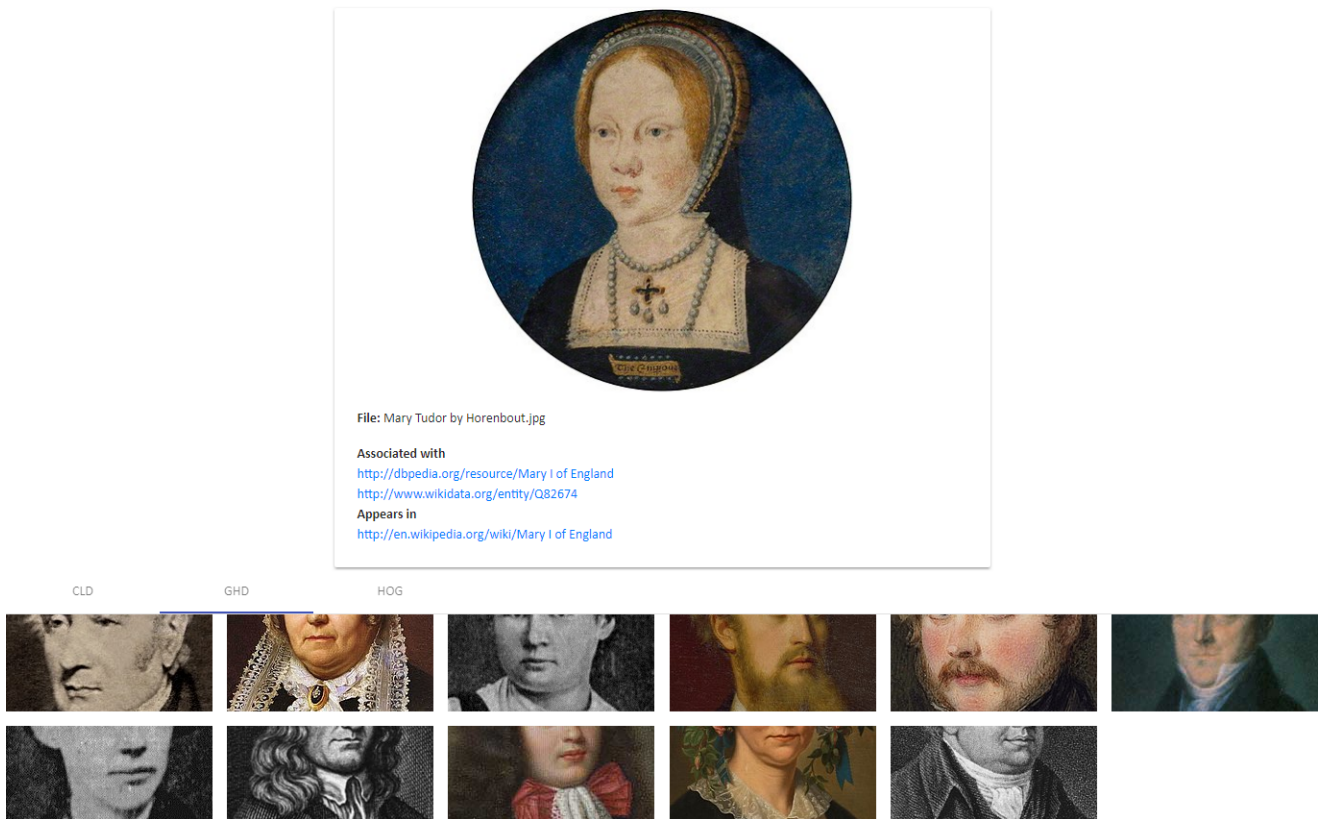


Figure 2: Detail page for visual entities; URL displayed: http://imgpedia.dcc.uchile.cl/detail/Mary_Tudor_by_Horenbout.jpg

there are many features that can be added, many use-cases that can be conceived, and many research questions that may arise from the current version of IMGPEDIA in the intersection of both areas.

The following are some of the directions we envisage:

Novel visual descriptors: Currently IMGPEDIA visual relations are based on three core descriptors relating to colour, intensity and edges. In some cases these descriptors can give good results, while in others the results leave room for improvement. An interesting direction to improve the similarity relations in IMGPEDIA is to incorporate and test novel visual descriptors and potentially combinations thereof. However two major challenges here are scale and evaluation. For scale, computing similarity joins over 15 million images requires specialised (and approximated) methods to avoid a quadratic behaviour; furthermore, for some descriptors even computing the descriptors is computationally challenging (for example, we performed initial experiments with a neural-network-based descriptor DeCAF [3] but estimated it would take years to compute over all images on our hardware). Another challenge is evaluation, which would require human judgement but where creating a gold-standard *a priori* is unfeasible; while human users could help to estimate precision, estimating recall seems a particular challenge.

Specialising resource links: As it was discussed previously with the example of Jean Auguste Ingres, our links from images to related knowledge-base entities is rather coarse being based on the existing image relations in external knowledge-bases, and the appearance of an image in the associated Wikipedia article of that image. It would thus be interesting to enhance IMGPEDIA by offering more specialised types of links between images and resources, such as to deduce (with high confidence) that a given entity is really depicted by an image. More ambitiously, it would be interesting to use Media Fragments [9] to try to identify which parts of the image depict a given entity.

Image-based neural networks: There exist a variety of neural networks trained over millions of images to identify different types of entities in images (e.g., [5]). Given that our images are linked to external knowledge-base entities, we could use these links to create a very large labelled dataset of images where labels can take varying levels of granularity (e.g. latin american government palaces, government palaces, government buildings, buildings, etc.). Such a dataset could be used to train a novel neural network and/or to evaluate current machine vision techniques over labelled images of varying levels of granularity. Furthermore, neural



Figure 3: Government palaces in Latin America

networks could enable us to describe what is happening in the image by extracting triples about the scene [1], to provide deep-learning based feature vectors [3]; to verify that an image is indeed a depiction of a linked entity; and so forth.

Image-based reasoning: We are planning to add new relations between images, such as `near_copy` or `same_subject`; potentially such relations may allow to perform novel types of reasoning over the data; for example if two entities in an external knowledge-base are related to a pair of images that are related by `near_copy` or `same_subject`, we could deduce new relations about the entities: if both entities are people, we can say that they have met (with some likelihood); or if one entity is a person and the other a place, there is a high chance that the person has visited the place.

SPARQL similarity queries: Extending SPARQL with functions to dynamically compute similarities between images after a semantic filter is applied, would obviate the need to rely on a bounded number of statically computed relations. Doing so efficiently may require hybrid cost models that understand the selectivity of not only the data-centred results, but also the similarity-based results. We are thus starting to research different ways of indexing the data to efficiently perform similarity joins.

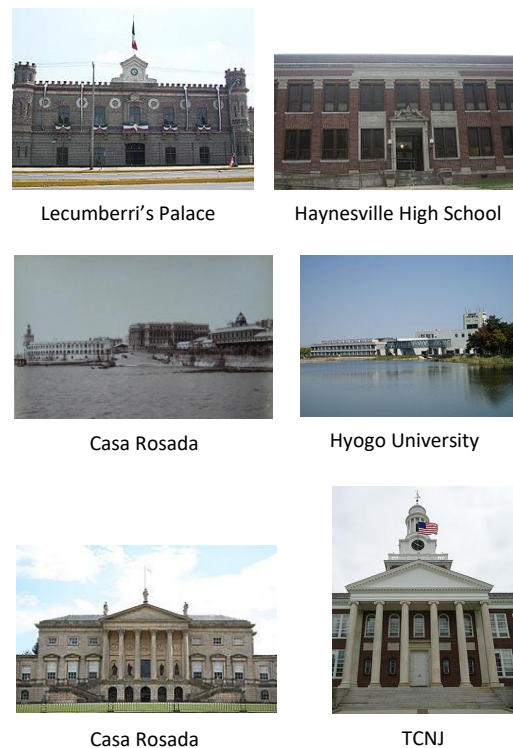


Figure 4: Educational institutions similar to Latin American government palaces

Video and other multimedia: Though WIKIMEDIA resources are predominantly images, there are also resources relating to video, audio, etc. Although audio would require a very different type of processing, it may be possible in future to try to match stills of video to images in IMGPEdia to, for example, identify possible topics of the video, or in general to bootstrap a similar form of semantic retrieval for videos. Likewise other sources of multimedia (e.g., YouTube, flickr, etc.) could be linked to IMGPEdia in the future.

Updating IMGPEdia: Currently the initial version of IMGPEdia is built statically. An interesting technical challenge then is to maintain the knowledge-base up-to-date with WIKIMEDIA and the external knowledge-bases, including, for example, the ability to incrementally update similarity relations.

While the prior topics relate to extending or enhancing IMGPEdia, an orthogonal but crucial aspect is the development of user applications on top of the knowledge-base. While the interface described here is a step in that direction – allowing users to explore the images, their associated resources, and their similarity relations – there is potentially much more left to be done. As a first step, we would like users to be able to specify keyword searches over the images. More ambitiously, for example, we would like users to be able to pose more complex visuo-semantic queries over the data (currently this still requires knowledge of RDF and SPARQL).

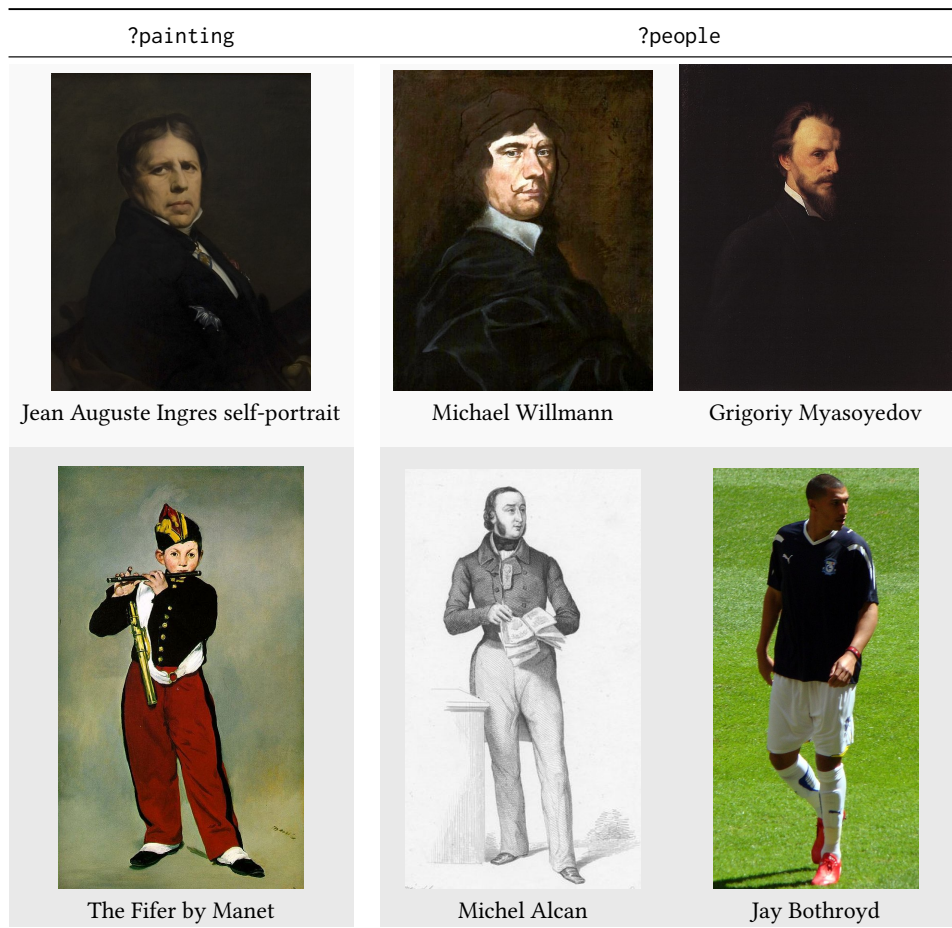


Figure 5: People similar to paintings in the Louvre



Figure 6: Paintings similar to other paintings at the Louvre; the image on the right appears in the Olympia artwork article while the (near-copy) image on the left appears in the article of its painter, Édouard Manet.

Such development of applications on top of IMGpEDIA would also require appropriate usability testing to validate. In general, we believe that through its focus on multimedia and use of existing knowledge-bases, IMGpEDIA has the potential to become a tangible demonstration of the value of Semantic Web technologies.

Another important question that we wish to address is: *how can IMGpEDIA compliment the existing projects of the Wikimedia Foundation?* Could the similarity relations computed by IMGpEDIA be added, for example, to the associated descriptions in WIKIDATA?

Could the detection of near-duplicate images help in some way the editors of WIKIPEDIA? Could some further visual information be added back to the WIKIMEDIA pages for images? What other possible use-cases might there be for a knowledge-base such as IMGpEDIA in the context of the varied WIKIMEDIA projects? We are eager to discuss and explore such use-cases.

Acknowledgements. This work was supported by the Millennium Institute for the Foundations of Data, by CONICYT-PFCHA 2017-21170616 and by Fondecyt Grant No. 1181896. We would like to thank Larry González and Camila Faúndez for their assistance.

REFERENCES

- [1] Stephan Baier, Yunpu Ma, and Volker Tresp. 2017. Improving visual relationship detection using semantic modeling of scene descriptions. In *International Semantic Web Conference*. Springer, 53–68.
- [2] Sean Bechhofer, Kevin Page, David M Weigl, György Fazekas, and Thomas Wilmering. 2017. Linked Data Publication of Live Music Archives and Analyses. In *International Semantic Web Conference*. Springer, 29–37.
- [3] Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell. 2014. Decaf: A deep convolutional activation feature for generic visual recognition. In *International conference on machine learning*. 647–655.

- [4] Sebastián Ferrada, Benjamin Bustos, and Aidan Hogan. 2017. IMGpedia: a linked dataset with content-based analysis of Wikimedia images. In *International Semantic Web Conference*. Springer, 84–93.
- [5] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems 25*. Curran Associates, Inc., 1097–1105.
- [6] Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick Van Kleef, Sören Auer, et al. 2015. DBpedia—a large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web* 6, 2 (2015), 167–195.
- [7] Albert Meroño-Peñuela, Rinke Hoekstra, Aldo Gangemi, Peter Bloem, Reinier de Valk, Bas Stringer, Berit Janssen, Victor de Boer, Alo Allik, Stefan Schlobach, et al. 2017. The MIDI Linked Data Cloud. In *International Semantic Web Conference*. Springer, 156–164.
- [8] Yves Raimond, Christopher Sutton, and Mark B. Sandler. 2009. Interlinking Music-Related Data on the Web. *IEEE MultiMedia* 16, 2 (2009), 52–63. <https://doi.org/10.1109/MMUL.2009.29>
- [9] Raphaël Troncy, Erik Mannens, Silvia Pfeiffer, and Davy Van Deursen. 2012. Media Fragments URI 1.0. W3C Recommendation. (2012).
- [10] Gaurav Vaidya, Dimitris Kontokostas, Magnus Knuth, Jens Lehmann, and Sebastian Hellmann. 2015. DBpedia Commons: Structured multimedia metadata from the Wikimedia Commons. In *International Semantic Web Conference*. Springer, 281–289.
- [11] Denny Vrandečić and Markus Krötzsch. 2014. Wikidata: a free collaborative knowledgebase. *Commun. ACM* 57, 10 (2014), 78–85.

A SPARQL QUERIES

Here we provide the SPARQL queries used to generate the examples.

Listing 1: SPARQL Query for images of Latin American government palaces using a federated query to Wikidata

```
PREFIX imo: <http://imgpedia.dcc.uchile.cl/ontology#>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
PREFIX wd: <http://www.wikidata.org/entity/>

SELECT ?imgu ?name ?cname WHERE {
  SERVICE <https://query.wikidata.org/bigdata/namespace/wdq/sparql>{
    ?palace wdt:P31 wd:Q16831714 ; # type government palace
    wdt:P17 ?country .
    ?country wdt:P361 wd:Q12585 . # countries in Latin America
  }
  OPTIONAL{
    ?palace rdfs:label ?name .
```

```
?country rdfs:label ?cname
  FILTER(LANG(?name)='en' && LANG(?cname)='en')
}
}
?img imo:associatedWith ?palace ;
  imo:fileURL ?imgu .
}
```

Listing 2: SPARQL Query for images of educational centers similar to Latin American government palaces

```
SELECT DISTINCT ?img ?img2 ?name ?label WHERE{
  SERVICE <https://query.wikidata.org/bigdata/namespace/wdq/sparql>{
    ... # obtain government palaces per Listing 1
  }
  ?img imo:associatedWith ?palace ;
    imo:similar ?img2 .
  ?img2 imo:associatedWith ?wiki .
  FILTER(CONTAINS(STR(?wiki), 'wikidata.org'))
  SERVICE <https://query.wikidata.org/bigdata/namespace/wdq/sparql>{
    ?wiki wdt:P31/wdt:P279* wd:Q2385804; # subclass of edu. institute
    rdfs:label ?label .
    FILTER(LANG(?label)='en')
  }
}
```

Listing 3: SPARQL query for retrieving people similar to paintings from the Louvre

```
SELECT ?painting ?people WHERE{
  SERVICE <https://query.wikidata.org/bigdata/namespace/wdq/sparql>{
    ?paintingw wdt:P31 wd:Q3305213 ;
    wdt:P276 wd:Q19675 ;
    rdfs:label ?label .
    FILTER(LANG(?label)='en')
  }
  ?paintingw imo:associatedWith ?paintingw ;
    imo:similar ?people .
  ?people imo:associatedWith ?peoplew .
  FILTER(CONTAINS(STR(?peoplew), 'wikidata.org'))
  SERVICE <https://query.wikidata.org/bigdata/namespace/wdq/sparql>{
    ?peoplew wdt:P31 wd:Q5 .
  }
}
```