Never ending face recognition for the LoD Cloud

Julien Subercaze, Christophe Gravier

Never ending face recognition for the LoD Cloud

Julien Subercaze
Laboratoire LT2C, Télécom Saint-Etienne
Université Jean Monnet
25 rue du Dr Remy Annino
F-42000 Saint-Etienne
Email: julien.subercaze@univ-st-etienne.fr

Christophe Gravier
Laboratoire LT2C, Télécom Saint-Etienne
Université Jean Monnet
25 rue du Dr Remy Annino
F-42000 Saint-Etienne
Email: christophe.gravier@univ-st-etienne.fr

***Abstract*—In this paper, we present FoP, a never ending face recognition learner on the Web. The first iteration of the face recognition model pro- pelling FoP was trained using Freebase data about politicians and their pictures. FoP is a never-ending system: when a new face is recognized, the learned model is updated ac- cordingly. At this step, FoP is also giving data in return to the LoD cloud that fed him in the first place: it leverages Linked Data on people recognized in these pictures, and on which rectangle area. Provided that both machine learning models and the LoD cloud are continuously struggling for more data for refinement processes, we demonstrate how to take advantage of their respective advantages to publish and to link more data on multimedia fragments on the Web.**

I. Introduction

The LOD-Cloud was first feeded by manually generated ontologies, such as OpenCyc, Freebase, ... This approach, while providing a good quality of data, fails to scale for the Web. One of the main difficulty was the requirements on the user to be able to contribute to such projects. Other limitations include the narrow range of the domains modeled by those ontologies. Meanwhile, wikis offered a very convenient framework for sharing general knowledge in an unstructured manner.

As a consequence, many projects focused on the automated leverage of knowledge from this unstructured data. Wikipedia being the most popular wiki, several projects used for enriching the LOD-Cloud with new structured knowledge.

This approch suffers two main drawbacks. Firstly it is prone to human mistakes when editing wikis, or spam (because of free-riding). Secondly, this approach does not take into account the large amount of knowledge that is already present on the Web.

The third generation approach aims at tackling this issue by walking the entire Web corpus and therefore learning knowledge at web scale.

However, the size of the Web makes it difficult to leverage knowledge from all this corpus in a single shot. To overcome this issue, incremental learning approaches have demonstrated significant improvement for the task of automated integration of textual data. This principle was popularized by NELL, the never-ending learner [1], that has been running continuously, and has learnt more than 50 millions of facts by itself. Recently, NELL was integrated into the LOD-Cloud [**?**].

These approaches primarily leverage textual data to the LOD-Cloud. Several initiatives have been engaged to tackle the issue of model multimedia data for the LOD-Cloud [2], [3]. Currently, multimedia data in the LOD-Cloud appears in the form of *related* to a given concept, with projects like FlicWrapper and DBPedia. Because they are human generated, they are prone to the same weaknesses as textual approaches (errors, spam). They also suffer from the lack of scalability : aforementioned services represent a tiny portion of available images on the Web. Large scale integration of multimedia data remains challenging. This is due to very nature of multimedia data. Moreover, a picture can contain multiple entities from a model. For instance a picture may depict three sports car of different brands. One may want to query the data about the relative positions of these cars or about the brands of the cars.

In this paper, we are interested in a particular sort of multimedia data: the presence of people in the image. We present our system `FoP` that is able to continuously learn new representation of people in a picture. In a first part `FoP`has to be bootstrapped using labelled data. For this, `FoP` needs a dataset with structured data already linking people resources and images in which this resource is present. The LoD-cloud provides such datasets, and using them to bootstrap `FoP` allows to circumvent the usual pitfall that only he who has the data can learn the model. In a second part, `FoP` learns new visual knowledge from the Web and makes its result available into the LoD-Cloud. The main contributions of the paper are the following :

- we propose a never-ending learner that is able to assess the presence of a person in a picture and to specify where the person is in the picture.
- We show that a very low error rate is possible : `FoP` achieves an error rate under 7%
- `FoP` can be queried online by both machines and humans. We provide a new data source into the LoD-Cloud, available through a SPARQL endpoint.
- We demonstrate the virtuous circle between the LoD cloud and a visual knowledge learner: the more the visual knowledge learner is fed by LoD data, the more it is accurate hence the more it gives data back to the LoD cloud.

The paper is organized as follows : Section **??** presents the bootstrapping part of our system. Section **??** describes the general architecture and the behaviour of the learner. Section **??** presents experimental results. Section **??** describes usage of the public online service. Finally Section **??** concludes.

II. Related Work

`FoP` aims at providing a fine-grained knowledge representation in the Semantic Web informational space about the presence of people in pictures. Given the tremendous amount of photos on the Web due to the ever-increasing multimedia consumption in our society, this is a problem that has gained interest over the recent years. We classified related works into two major categories. The first one encompasses approaches in which the knowledge representation is provided by a crowdsourcing mechanisms, which is not uncommon in linked data management tasks [4]. At the opposite, `FoP` falls into the second categories that is automatic person recognition. Both categories exhibit complementary strengths and shortcomings. A manual process, even crowdsourced, may not scale as well as an autonomic approach. Nonetheless, humans may encode knowledge that is difficult for the machine to learn, such as the relative position of persons in the picture, the fact that a person is a sunglass bearer, etc. That is the reason why this category is interesting as related works in order to have an idea on how `FoP` performs in terms of knowledge representation.

At the opposite to crowdsourced solutions, an automatic process may not be able to grasp all the latent knowledge in the picture about people - or to grasp it with precision - yet it may well tend to address Web scale. Crowdsourced and automatic processes are a trade-off between precision and recall. In this section, we will review related works that falls into these two categories as follows. On one hand, crowdsourcing approaches will be reported and analyzed with a prority given to how fine-grained is the knowledge representation. On the other hand, we will be primarily interested in the scalability and precision of automatic person recognition processes. Both will be discussed in term of data lift abilities. In Table I, we provide a synoptic view on the related works reported in this section.

Crowdsourcing person recognition and data lift for multimedia Web resources. The Flickr wrapper [5] intends to extend Wikipedia with RDF annotations on Flickr pictures. Any Web user is invited to declare links between a Flickr photo and a Wikipedia page. The linkage possibilities are only available at a coarse grain: a user only declares that a picture is "related" to a Wikipedia page (related can be pretty vague), and it is not possible to link only a subpart of the picture (e.g. faces). No control is applied on the crowsourced data, while free-riding is prone to error, and no metric aside human validation on such a large corpus is made available. With the assumption that video objects are an opaque information nutshell for crawlers, [6] presents a generic crowdsourcing framework for automatic and scalable event detection framework for HTML 5 videos, especially for YouTube. This framework is easing the leveraging of Linked Data based on the Event Ontology[7], which includes the `Agent` class whose instances can be persons. It partially relies on the end-user behaviour when reading the video in order to detect events. The authors state that the precision of the inferred knowledge from these observations improve because of the wisdom of the crowd. However, no evaluation of this improvement, neither its absolute performance to a ground truth is provided to support this claim. In [7], the authors explore the utility of tasks marketplaces like Amazon Mechanical Turk [14] (AMT). In this project, the Human Intelligence Task (HIT as denoted in AMT) is to identify figure ground masks for different object categories on the PASCAL VOC 2010 dataset [15]. The interface of the task allows to outline polygons in images in order to draw the outline of recognized object by the human performing the HIT. Among all objects in the PASCAL dataset, 7296 are persons (31.2%). The validation is performed against the PASCAL 2010 dataset ground truth. While the model does not evolve with time and does not provide a data lift mechanism, it however provides a leveraging of fine-grained information on people in a picture, especially gender, race, age, hair-type, glasses, shoes, lower-clothes, ... This work is also interesting because it enforces the assumption stated at the begining of this section: the recognition task performed by humans are pretty good since only 10% of submitted boundaries were rejected, under tight validation constraints. Photocopain [8] is a semi-automatic annotation system. It tries to ease the burden of photo annotation from its end-users, while using human annotation as a very high confidence datasource. The main application is archiving the personal experience of its user. Among several features, it provides a person detection, recognition and annotation service. The face detection model is built using a hue, intensity, and texture map, whose precision is tuned by correlating its output with the EXIF data of the photo. The result is stored in a RDF triplestore, and proposed to the end user for validation and metadata expansion (including person tagging and recognition). Tested on a limited dataset of 150 images including a 30 images tests set, Photocopain exhibits promising result for photo annotation.

Automatic person recognition and data lift. In [9], the authors present a life-logging system that combines users' private log on their smartphone and data obtained automatically from the Web. The face recognition system is mainly performed using Bluetooth scans of nearby devices, whose owner had been previously declared in the system database. The user's picture are then lifted to the LoD-cloud mainly using the `foaf` ontology which is limited vocabuluary if solely used, and does not allow an excellent precision. Nonetheless, this can be explained by the fact that the application not only address person recognition, but also other personal events that the end user wants to reccord in his personal life-blogging system. In [10], the authors present Pharos, an European initiative to index and search audiovisual contents. In Pharos, a Face Annotation service submodule is responsible for indexing persons which are depicted in a picture. It relies on a previously learned model, on an unknown dataset. This submodule is not reported to embed online learning mechanisms. This module is not exploiting neither publishing Linked Data, but instead focused on publishing MPEG7 compliant metadata. Last year [11] presented FANS, a face annotation framework. The

[1]Ranging from $--$, $-$, $+$, $++$ for respectively personal data, small datasets (in hundred or thousands), large dataset (hundred of thousaunds), Web scale). `N.S.` stands for "not specified".

[2]Ranging from $--$, $-$, $+$, $++$ for respectively no person identification, linkage of person "related to" a resource, linkage of region of picture with a resource, fine-grained information such as person position, hair colour, etc.

[3]The baseline in this case was a learnt model using a face recognition algorithm.

[4]Relates to the ability of the system to reinfine its person recognition mechanisms beyond the learnt model.

[5]w.r.t. dbpedia updates

[6]*foaf* only

[7]http://motools.sourceforge.net/event/event.html

	Related works								
System name	Flickrwrappr	N/A	N/A	Photocopain	Imouto	Pharos	FANS	N/A	NEIL
Reference	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]
Face recognition	crowdsourced	crowdsourced	crowdsourced	semi-auto	auto	auto	auto	auto	auto
Working scale[1]	−	−	−	−−	−−	N.S.	++	+	++
Knowledge granularity[2]	−	−	+	−	−	−	−	−	−−
Data validation	none	none	test set	test set	baseline[3]	end-users	test set	none	manually
Model evoluates w/ time[4]	no	no	no	no	no	no	no	yes[5]	yes
Images with 2+ persons	yes	yes	yes	yes	no	no	no	yes	no
Linking to the LOD cloud	yes[6]	yes	no	yes	yes[6]	no	no	yes[6]	no

TABLE I. ANALYSIS OF RELATED WORKS FOR LEVERAGING MULTIMEDIA/PERSON METADATA OUT OF PICTURES.

FANS inner model is first trained using a manually constructed dataset resulting from google queries on 6,025 persons. By applying a Locality-Sensitive Hash function to the natural feature points of detected faces, the authors demonstrated a scalable Web image retrieval engine. However, FANS do not interact with the Web of Data, and the model is not evolving with time. Another work provided by [12] stresses the usage of *sponging* multimedia information, as intended in `FoP`. *Sponging* is the task to leverage new triples from Web crawls and large textual corpus. The authors propose to search `dbpedia` for the persons depicted in a resource (a `foaf:image`). The authors assume that two or more persons are present in the picture, and that face detection and face recognition algorithms can fail to identify all persons in a given picture. If at least one person is devised, the sponging process is started : the system looks up for related persons in `dbpedia` for the recognized persons. A new iteration of face recognition is then started, limited to the `foaf:depicts` properties matching related persons. The shortcomings of this approach is that the picture must contain at least two persons, and that all these persons can be recognized and linked in DBPedia. Lately [13] proposed NEIL, a Never Ending Image Learner. Like `FoP`, NEIL is a never-ending learner that aims at extracting visual knowledge from various images. NEIL discovers terminology and instance classification information. It is a large scale initiative in term of CPU consumption (more than 350K CPU hours) for processing the 2 millions images downloaded by NEIL (and counting). NEIL is bootstraped using text-based indexing tools, primarily Google Image Search. Among the 1152 manually entered or discovered object categories in NEIL, some include visual knowledge for people. This includes learning that eye is a part of Baby, for instance. However, it does not provide instance labelling to existing person. Instead, NEIL is interested in discovering the visual descriptor of a given concept. For instance, NEIL discovered the Actor object class[8]. Among classified photo in this object class, there are indeed mainly actors that have been correctly classified as such by NEIL. There is no possibility however to link the image with an actor, neither to know which actor is depicted in the image. NEIL does not support multiple person in an image, neither does it provides a sponging process from its database to the LOD cloud.

III. FoP OVERVIEW

Our system, called `FacesOfPolitics` is made of two steps. In a first part, detailled in the next section, we bootstrap the face recognition model using data from Freebase[9]. The aim of this step is to leverage existing LOD-Cloud data to avoid as most as possible classical cold start issues.

In a second step, which is the steady state of `FoP` , our system crawls RSS feeds of major news websites, looking for new pictures. Whenever a face is detected and matched to a known person, the internal triplestore stores this information (See Figure **??**). Following the idea of never-ending learner that has been succesfully applied to language learning [16], our model is being updated and improved whenever a person has been recognized in a new picture. This step is detailled in section **??**

An overview of this global architecture is depicted in figure **??**.

INCLUDE CG PICTURE HERE

IV. BOOTSTRAPPING

`FoP`'s central component is the face recognition model that is used both to identify persons in images and to be updated throughout the never ending learning process. The problem of initiating the model is twofold. In the first place, the algorithm behind the face recognition must be efficient, i.e. in our case provide a low false positive rate and it must also be rapidly updated whenever new pictures arise. In the second place, the model has to be initiated with reliable data in a sufficient volume.

The Local Binary Patterns algorithm [17] is a very robust face recognition algorithm that is less sensitive to lighting conditions than other standard algorithms. Moreover it is not computationally expensive to perform an update of the model, which is a unique feature among holistic methods [18].

Face recognition algorithms are very sensitive to paramaters settings. For this purpose, we conducted several tests to determine the best parameter set. Experiments conducted on a sample dataset showed that it is not possible to maintain high values for both precision and recall, as depicted in Figure 2. This is due to the fact that the task of face recognition is still a hard research issue. As our system will handle a large volume of pictures, and that our goal is to offer the correct data, we decided to trade recall for precision.

We demonstrate our approach on recognizing French politicians in national news. Numerous celebrities are identified in the LOD-Cloud. We extract the list of living politicians having government positions in France along to their pictures.

[8]http://ladoga.graphics.cs.cmu.edu/xinleic/NEILWeb/content.php?category=Objects\&class=actor

[9]http://www.freebase.com

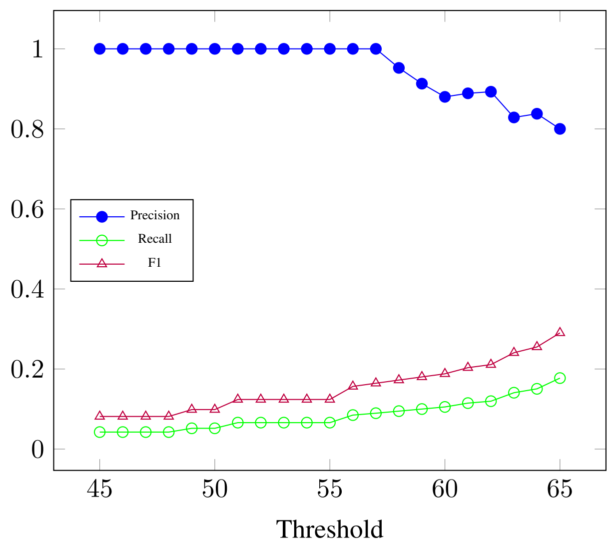

Fig. 1. Face recognition : influence of Binary Local Pattern parameter on precision and recall

For each person, the number of pictures ranges from one to three. We then use these pictures and their associated label (i.e. the person identifier) to train a first version of the face recognition model. For this first model, as depicted in Figure 2 on the iteration zero, we chose a threshold of 57 (which is the maximum authorized distance between two pictures to be considered neighbours), that maximizes the recall for a perfect precision. However the recall still remains low and very few pictures would be handled.

In order to continuously provide pictures and to enhance our model, we crawl several RSS feeds dealing with political news. For each article we extract pictures, detect faces and match them against the previously trained model. If a person is detected using the face recognition model, we validate its presence by searching the article for his/her name. If validated, data is added to the internal database (see next section), and the model is updated using the newly detected face.

Updating the model allows us to overcome the initial low recall issue. As shown in Figure 2, for a threshold of 57, we maintain a very high precision (above .93). In the mean time, recall greatly improved from .10 to .35. For greater values of the threshold the recall improvement is far greater but it implies serious degradation of the precision to unacceptable values, that would result in numerous falsely labeled pictures.

V. FoP Architecture

The `FoP` system data leveraging flow is illustrated in Figure 3. The exchange of data between `FoP` and the LoD-cloud is giving rise to raising edges of data called "`ticks`", while the falling edges (from the LoD-cloud to `FoP`) are respectively called "`tocks`". These two distinct percolations of data are of uneven complexity. In what follows we first depict `tocks` phases, as being a more straightforward process and which initiate the `FoP` never-ending learner. We then present `ticks` occurring in `FoP`.

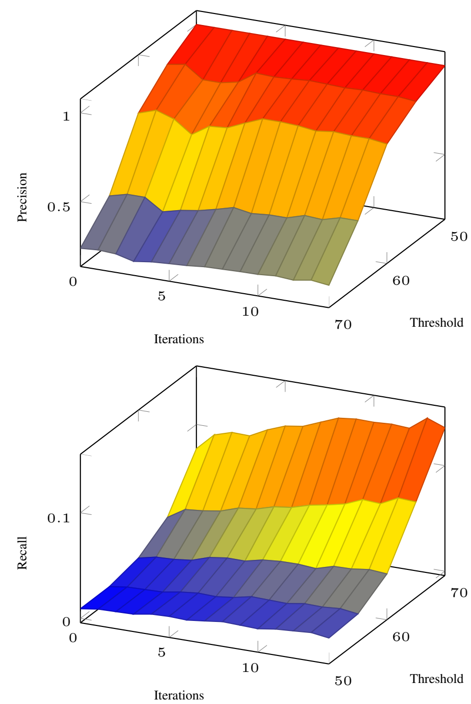

Fig. 2. Influence of updates on precision and recall

A. Tocks

`Tocks` are data queried from the LoD-cloud that enrich the `FoP` learner. A first `tock` occurred when we initialized `FoP`. The first `tock` drived the training of a first face recognition model. For this, the Local Binary Patterns algorithm [19] is a very robust face recognition algorithm that is less sensitive to lighting conditions than other standard algorithms. Moreover it is not computationally expensive to perform an update of the model, which is a unique feature among holistic methods [20]. Face recognition algorithms are however very sensitive to parameters settings. For this purpose, we conducted several tests to determine the best parameter set to use at `tock` phase. Experiments conducted on a large sample dataset showed that it is not possible to maintain high values for both precision and recall, as depicted in Figure **??**.

This is due to the fact that the task of face recognition is still a hard research issue. As our system will handle a large volume of pictures, and that our goal is to provide data with the highest level of correctness, we trade recall for precision. In order to determine the best parameter that maximize precision while preserving recall, we learnt a first model with French politicians in national news. Numerous celebrities are identified in the LOD-Cloud. We extract the list of living politicians having government positions in France along to their pictures in Freebase. For each person, the number of pictures ranges from one to three. We then use these pictures and their associated resource to train a first version of the face recognition model. For this first model, we experimentally set the single parameter (which is the maximum authorized distance between two pictures to be considered neighbors) of the Binary Local Pattern algorithm to 57. As illustrated in Figure 1, this value maximizes the recall for a near-perfect precision. However, the recall still remains low and very few pictures would be handled. That is the reason why other `tocks` occur at regular interval of time. At further `tocks`, `FoP` queries again the Freebase database for new Picture–Person links. For each of the new pictures that were introduced in Freebase since the last query, `FoP` updates the learnt models by introducing whether new reference images for a given person, or by introducing a new person along the image(s) in which he/she appears in. At each `tock`, the `FoP` face recognition model is therefore enhanced using LoD-based data. The LoD based data are enhanced at each `tick`.

B. Ticks

A `tick` is initiated by receiving an update of one of the RSS feeds to which `FoP` has subscribed. For each article `FoP` extracts pictures, detects faces, and matches them against the previously trained model. If a person is detected using the face recognition model, we validate its presence by searching the article text for his/her name. If validated, annotated feeds articles are added to the internal database, and the model is updated using the newly detected face. A `tick` can therefore also generate model updates of the `FoP` recognition model. Updating the model allows us to overcome the initial low recall issue. As shown in Figure 1, for a threshold of 57, we maintain a very high precision (above .93).

In the mean time, recall greatly improved from .10 to .35. For greater values of the threshold the recall improvement is far greater but it implies serious degradation of the precision to unacceptable values, that would result in numerous falsely labeled pictures. Aside from updating the model and storing annotated feeds articles for later consultation by `FoP` users, `FoP` leverages Linked Data at `ticks`. The ontology for media resources (`ma-ont`[10]) defines media fragment as sub-parts of a media. In our purpose we use the relationship `hasFragment` to identify the relationship between the original images and the faces it contains. Mediafragments[11] makes it possible to specify rectangular clipping of images by appending its coordinates to the original URI. This is particularly useful for identifying several persons in a picture, since one can specify the region where the face of a person is located. Due to domain coverage shortcomings of `ma-ont`, we extended the ontology to be able to identify persons within an image. `ma-ont` has a property called `features` that is to be used for actor in movies. In our case, we intend to define that a person is present in picture, not acting. We therefore defined an object property `IsInPicture` whose domain is a `foaf:Person` and whose range is the intersection of `ma-ont#Image` and `ma-ont#MediaFragment`. This enables, among other, to search for pictures containing multiple people. For example one is now able to retrieve in one query pictures containing both France's president and prime minister.

VI. Providing Linked Data

The ontology for media resources (`ma-ont`) [12] defines media fragment as subparts of a media. In our purpose we use the relationship `hasFragment` to identify the relationship between the original images and the faces it contains. Mediafragments[13]makes it possible to specify rectangular clipping of images by appending its coordinates to the original URI. This is particularly useful for identifying several persons in a picture, since one can specify the region where the face of a person is located.

Due to limitations of `ma-ont`, we extended the ontology to be able to identify persons within an image. `ma-ont` has a property called `features` that is to be used for actor in movies. In our case, we intend to define that a person is present in picture, not acting. We therefore defined an object property `IsInPicture` whose domain is a `foaf:Person` and whose range is the intersection of `ma-ont#Image` and `ma-ont#MediaFragment`. This enables, among other, to search for pictures containing multiple people.

A public SPARQL endpoint links our dataset to the LOD. Full RDF Dumps are also available from the previously specified URL.

VII. Using FoP

Our `FoP` never ending learner is available online at http://demo-satin.univ-st-etienne.fr/facesofpolitics/. Figure 4 shows the current trained face recognition model fueling `FoP`. In

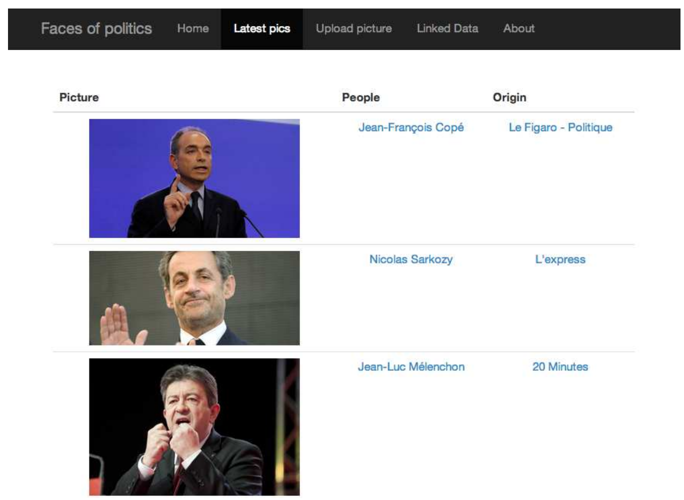

Fig. 4. Primary user interface for reading enriched RSS articles from subscribed feeds.

this demonstration, `FoP` is subscribing to the politicians news feeds of the New York Times, the BBC, Le Monde, Liberation, and El País. Using these settings, we want to demonstrate the benefits of never-ending learning from Linked Data and never-ending lifting new Linked Data in return. Especially, we will showcase the following to the conference attendees[14].

Browsing latest annotated feeds. Through the `FoP` user interface we shall demonstrate `FoP` in action by accessing latest feeds with recognized person(s). We will present discovered persons along the part of the multimedia in which they were found, along the context of the original feed article.

[10]http://www.w3.org/TR/mediaont-10/
[11]http://www.w3.org/TR/media-frags/
[12]http://www.w3.org/TR/mediaont-10/
[13]http://www.w3.org/TR/media-frags/
[14]A video is accompanying this paper: http://vimeo.com/81178843

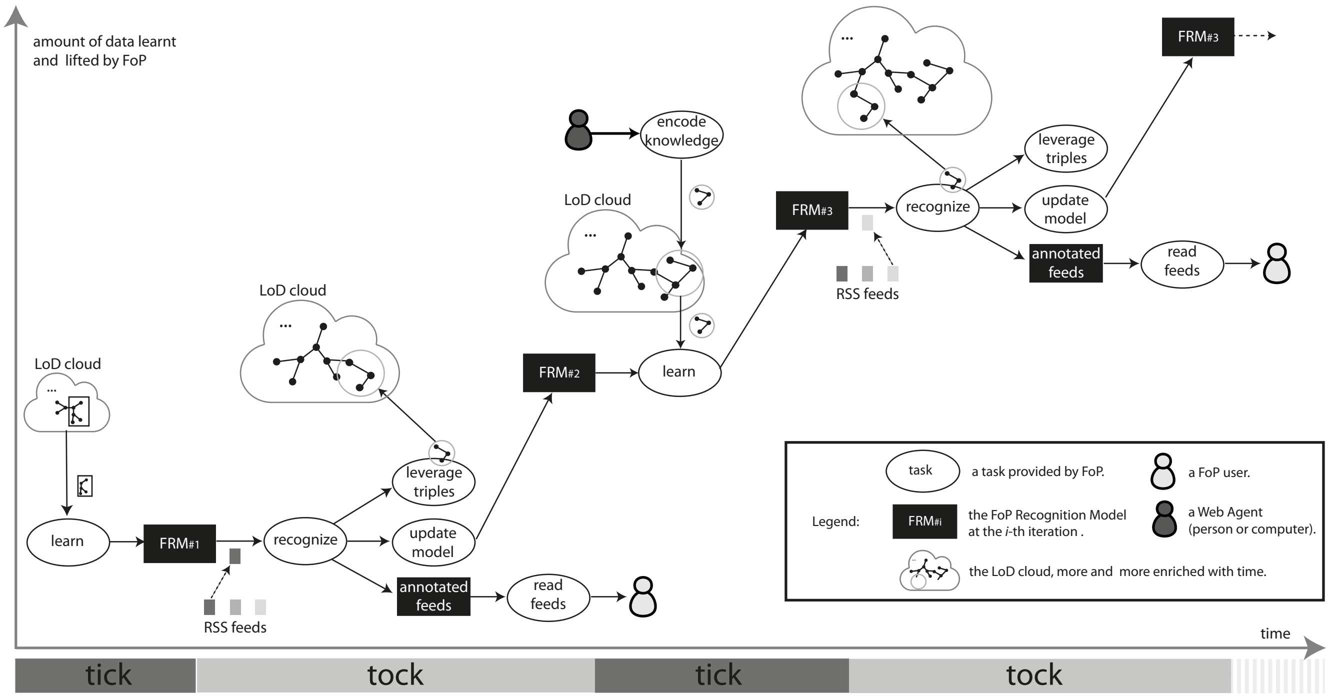

Fig. 3. Never-ending face recognizing and semantizing tick/tock model.

Article curation method on single or multiple person presence selection. By switching to another GUI area, one can query `FoP` data using SPARQL. This encompasses article retrieval based on the presence of one known person in the `FoP` triplestore. It is also possible to create a curation of articles from different RSS feeds by specifying several persons that must be present in the article illustration. Let us illustrate in details the fine-grained queries that `FoP` offers. One can query `FoP` for all pictures and associated articles in which one person is placed left to another in the picture. For example, let us query `FoP` for any picture in which Jean-Marc Ayrault is depicted left to François Hollande in all pictures in which `FoP` recognized both of them. This can be done using the following SPARQL query[15] : The result of the query is depicted at Figure 6 as taken from one major French media website[16].

Finally, we also provide a public SPARQL endpoint linking the Linked Data leveraged by `FoP` to LOD-cloud resources. Full RDF Dumps are also available from the previously specified URL[17].

```
SELECT ?media
WHERE {
 dbpedia:Jean-Marc_Ayrault fop:isInPictu
                        re ?medfragJMA .
 dbpedia:Fran%C3%A7ois_Hollande fop:isIn
                    Picture ?medfragFH .
 ?media maont:hasFragment ?medfragJMA  .
 ?media maont:hasFragment ?medfragFH   .
 BIND(str(?medfragJMA) as ?newJMAmedia).
 BIND(str(?medfragFH) as ?newFHmedia)  .
 BIND(STRBEFORE(STRAFTER(?newJMAmedia,
      '?xywh=pixel:'), ",") AS ?JMAabs)
 BIND(STRBEFORE(STRAFTER(?newFHmedia,
       '?xywh=pixel:'), ",") AS ?FHabs)
 FILTER (xsd:decimal(?JMAabs) <
                    xsd:decimal(?FHabs))
}
```

Fig. 5. SPARQL query for all pictures with Jean-Marc Ayrault positionned left to François Hollande.

VIII. Conclusion

`FoP` is a never-ending face recognition learner. Subscribing to several RSS feeds, it is able to recognize famous politicians, to auto-enhance its model and to serve the data as linked open data. Through a continuous virtuous circle between the LoD-cloud and `FoP` recognition model, `FoP` is leveraging annotated faces as Linked Data.

In the future, we plan to study the evolution of `FoP` over time, especially the evolution of its precision and recall as it scale for more faces, foreign politician but also other applications like sport as well. In this paper, we presented an approach to deliver correctly annotated images to the LOD-Cloud. By using a face recognition model, bootstraped with LOD-Cloud data, we are able to ensure a high precision for the presence of a person in a picture. The never-ending learning approach allows us to overcome the initial low recall issue.

[15]Using the following prefixes: `PREFIX dbpedia:<http://dbpedia.org/resource/>`, `PREFIX fop:<http://demo-satin.telecom-st-etienne.fr/facesofpolitics/public/ontologies/pplinpic#>`, and `PREFIX maont: <http://www.w3.org/ns/ma-ont#>`

[16]http://www.europe1.fr/Politique/Chomage-la-methode-Coue-de-Hollande-et-Ayrault-1686635/

[17]http://demo-satin.telecom-st-etienne.fr/facesofpolitics/rdf

Fig. 6. Media retrieved by FoP using query provided by Figure 5

Finally, we provide a publicly available dataset of politicians pictures that is constantly growing and being enhanced. In future research, we will aim at generalizing our system to larger categories of persons.

References

[1] A. Carlson, J. Betteridge, B. Kisiel, B. Settles, E. R. Hruschka Jr, and T. M. Mitchell, "Toward an architecture for never-ending language learning." in *AAAI*, 2010.

[2] G. Stamou, J. van Ossenbruggen, J. Z. Pan, G. Schreiber, and J. R. Smith, "Multimedia annotations on the semantic web," *Multimedia, IEEE*, vol. 13, no. 1, pp. 86–90, 2006.

[3] R. Arndt, R. Troncy, S. Staab, L. Hardman, and M. Vacura, "Comm: designing a well-founded multimedia ontology for the web," in *The semantic web*. Springer, 2007, pp. 30–43.

[4] *Crowdsourcing tasks in Linked Data management*, Workshop on Consuming Linked Data COLD 2012, co-located with the 10th International Web Conference ISWC2011. -, November 2011.

[5] F. U. Berlin. (2009) Flickr wrappr, precise photo association. [Online]. Available: http://wifo5-03.informatik.uni-mannheim.de/flickrwrappr/

[6] T. Steiner, R. Verborgh, R. Van de Walle, M. Hausenblas, and J. G. Vallé s, "Crowdsourcing event detection in youtube video," in *Proceedings of the 1st workshop on detection, representation, and exploitation of events in the semantic web*, M. Van Erp, W. R. Van Hage, L. Hollink, A. Jameson, and R. l. Troncy, Eds., 2011, pp. 58–67.

[7] S. Maji, "Large scale image annotations on amazon mechanical turk," EECS Department, University of California, Berkeley, Tech. Rep. UCB/EECS-2011-79, Jul 2011. [Online]. Available: http://www.eecs.berkeley.edu/Pubs/TechRpts/2011/EECS-2011-79.html

[8] M. M. Tuffield, S. Harris, D. P. Dupplaw, A. Chakravarthy, C. Brewster, N. Gibbins, K. O'Hara, F. Ciravegna, D. Sleeman, N. R. Shadbolt, , and Y. Wilks, in *Proceedings of the First International Workshop on Semantic Web Annotations for Multimedia (SWAMM)*, Edinburgh, May, held as part of 15th World Wide Web Conference (22-26 May, 2006).

[9] A. Smith, K. O'Hara, and P. Lewis, "Visualising the past: Annotating a life with linked open data," in *Web Science Conference 2011*, June 2011, event Dates: June 14th-17th 2011. [Online]. Available: http://eprints.soton.ac.uk/272324/

[10] A. Bozzon, M. Brambilla, P. Fraternali, and P. Pigazzini, "Integration of a human face annotation technology in an audio-visual search engine platform," in *Proceedings of the 2010 ACM Symposium on Applied Computing*, ser. SAC '10. New York, NY, USA: ACM, 2010, pp. 839–843. [Online]. Available: http://doi.acm.org/10.1145/1774088.1774261

[11] S. C. H. Hoi, D. Wang, I. Y. Cheng, E. W. Lin, J. Zhu, Y. He, and C. Miao, "Fans: face annotation by searching large-scale web facial images," in *WWW (Companion Volume)*, 2013, pp. 317–320.

[12] R. Verborgh, D. V. Deursen, E. Mannens, C. Poppe, and R. V. de Walle, "Enabling context-aware multimedia annotation by a novel generic semantic problem-solving platform." *Multimedia Tools Applications*, no. 1, pp. 105–129.

[13] X. Chen, A. Shrivastava, and A. Gupta, "Neil: Extracting visual knowledge from web data," in *Proc. 14th International Conference on Computer Vision*, vol. 3, 2013.

[14] G. Paolacci, J. Chandler, and P. G. Ipeirotis, "Running experiments on amazon mechanical turk," *Judgment and Decision making*, vol. 5, no. 5, pp. 411–419, 2010.

[15] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, Jun. 2010.

[16] A. Carlson, J. Betteridge, B. Kisiel, B. Settles, E. R. Hruschka Jr, and T. M. Mitchell, "Toward an architecture for never-ending language learning." in *AAAI*, 2010.

[17] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 12, pp. 2037–2041, 2006.

[18] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Computing Surveys (CSUR)*, vol. 35, no. 4, pp. 399–458, 2003.

[19] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 12, pp. 2037–2041, 2006.

[20] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Computing Surveys (CSUR)*, vol. 35, no. 4, pp. 399–458, 2003.