FoP: Never-Ending Learner for Multimedia Knowledge Extraction

Julien Subercaze, Christophe Gravier

FoP: Never-ending learner for multimedia knowledge extraction.

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 our system Faces of Politics (henceforth FoP), that is able to continuously learn multimedia knowledge of Web multimedia resources about the presence of person(s) in a pictures and to leverage this knowledge to the Linked Open Data cloud (LOD-cloud). FoP promotes both scalability of the data lift process for this domain and a structured knowledge representation for complex queries. The system was bootstraped using Freebase data about politicians and their pictures, and we show that the model provides a good generalization with an error rate below 7%. Meantime, FoP not only relates a person to a multimedia resource, but it also detects and publishes metadata on the position of the person in the picture. Moreover, it supports the presence of several persons in the picture. 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. This allows fine-grained queries like creating a curation of documents in which a person is depicted relatively to another for instance. On a technical point-of-view, we also provide a Website for browsing FoP knowledge base as Web users, and we also offer a public SPARQL endpoint for robots or other Web applications.**

I. Introduction

The LOD-Cloud was first initiated using manually generated ontologies, such as OpenCyc [1] or Freebase [2]. This approach, while providing a good quality of data, fails to scale for the Web. One of the main difficulty is the requirements on the user to contribute to the knowledge modeling tasks. Another limitation is the incomplete domain coverage of the underlying ontologies. Meanwhile, wikis offered a very convenient framework for sharing general knowledge in an unstructured manner. Social web users therefore massively contributed to these platforms and classification tasks, especially Wikipedia [3]. As a consequence, many projects focused on the automated leverage of knowledge from these unstructured data. For example, Wikipedia being the most popular wiki, several projects used it for enriching the LOD-Cloud with new structured knowledge (e.g. [4], [5], [6]). Such approaches mainly suffers from two limitations. First, they are prone to human mistakes when editing wikis, or spam (because of free-riding). Secondly, they does not take into account the large amount of knowledge that is already present on the Web, and may not intend to model all existing or future domain knowledge.

As a consequence, other works explore the possibility to tackle these problems by walking the entire Web corpus and therefore learning knowledge at web scale. The size of the Web makes it difficult to leverage knowledge from a significant part of the Web 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 [7], that has been running continuously, and has learnt more than 50 millions of facts by itself. Recently, we also contributed to this task by making NELL compliant with semantic Web standards [8].

These approaches primarily leverage textual data to the LOD-Cloud. Less efforts were provided for multimedia data with respect to those brought on textual data. Nonetheless, pictures and videos are a first-class citizen on the Web, which drove several works towards leveraging multimedia data for the LOD-Cloud [9], [10] (See Section II for a more detailled state-of-the-art). The first issue with multimedia data is that currently only a high-level of information is leveraged to the LOD-cloud. This mainly means that multimedia data in the LOD-Cloud appears in the form of link from an URI associated from a Person to an URI associated to a picture. When these information comes from Wikis, there are prone to the same weaknesses than their textual equivalents (errors, spam, scalability, broad range of high level taxonomies *versus* narrow domain to model). There is also a concern of level of details in the knowledge modeling task of multimedia data, that are the leverage of multimedia fragment. 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, whether they are in the foreground or not, etc. Although the standards do exist in the LOD-cloud [11], there are yet to be a learner that will both improve the scalability of learnt multimedia knowledge, while proposing a fine-grained representation of multimedia knowledge.

In this paper we present our system FoP that is able to continuously learn from the Web new multimedia knowledge about the presence of person(s) in a pictures and to leverage this knowledge to the LOD-cloud. This allows queries on structured content, therefore providing a richer expressivity of the query. 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 of multimedia fragments 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.

On a technical point-of-view, we also provide a publicly available service compliant with semantic Web standards. The paper is organized as follows : Section II related works relevant to FoP. Section IV presents the bootstrapping part of our system. Section V describes the general architecture and the behaviour of the learner. Section VI describes usage of the public online service and provides some illustrations. Section VII 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 [12]. 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 [13] 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 sub-part 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, [14] presents a generic crowd-sourcing 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[1], 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 [15], the authors explore the utility of tasks marketplaces like Amazon Mechanical Turk [16] (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 [17]. 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 [18] 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 [19],

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

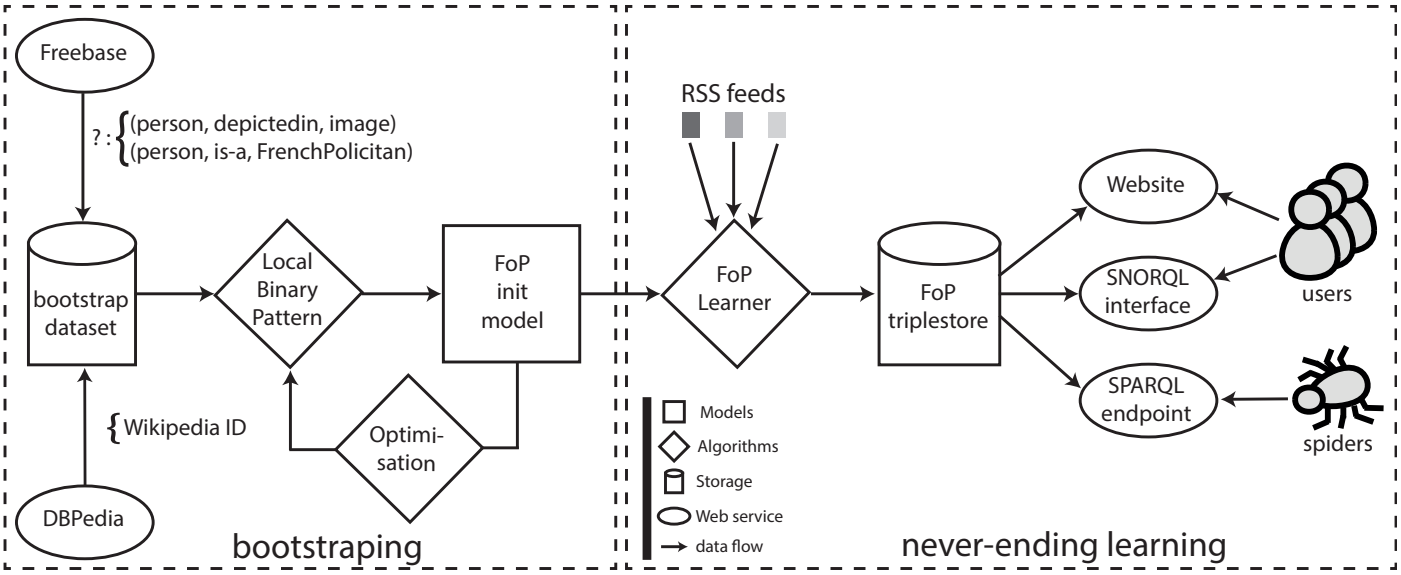

Fig. 1. FoP overview. Left part presents the bootstrapping step (i.e transient state) while the right part shows the steady state step of the system.

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 [20], 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 [21] presented FANS, a face annotation framework. The 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 [22] 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 [23] 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[2]. 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, detailed in the next section, we bootstrap the face recognition model using data from Freebase[3]. The aim of this step is to leverage existing LOD-Cloud data to avoid as most as possible classical cold start issues.

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

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

[4]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".

	Related works								
System name	Flickrwrappr	N/A	N/A	Photocopain	Imouto	Pharos	FANS	N/A	NEIL
Reference	[13]	[14]	[15]	[18]	[19]	[20]	[21]	[22]	[23]
Face recognition	crowdsourced	crowdsourced	crowdsourced	semi-auto	auto	auto	auto	auto	auto
Working scale[4]	−	−	−	−−	−−	N.S.	++	+	++
Knowledge granularity[5]	−	−	+	−	−	−	−	−	−−
Data validation	none	none	test set	test set	baseline[6]	end-users	test set	none	manually
Model evoluates w/ time[7]	no	no	no	no	no	no	no	yes[8]	yes
Images with 2+ persons	yes	yes	yes	yes	no	no	no	yes	no
Linking to the LOD cloud	yes[9]	yes	no	yes	yes[9]	no	no	yes[9]	no

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

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 1). Following the idea of never-ending learner that has been succesfully applied to language learning [24], our model is being updated and improved whenever a person has been recognized in a new picture. This step is detailled in section V. An overview of this global architecture is depicted in figure 1.

IV. BOOTSTRAPPING (TRANSIENT STATE))

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 powering the face recognition module 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.

There are several approaches to perform face recognition. It is however out of the scope of this paper to provide an extensive review of existing literature. The interested reader may referto the following surveys : [25], [26], [27], [28].

Nonetheless, the Local Binary Patterns algorithm [29] 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 [28]. It is therefore our algorithm of choice for FoP.

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 3. This is due to the fact that the task of face recognition is complex and is still a hard research issue. As our system will handle a large volume of pictures, and that our goal

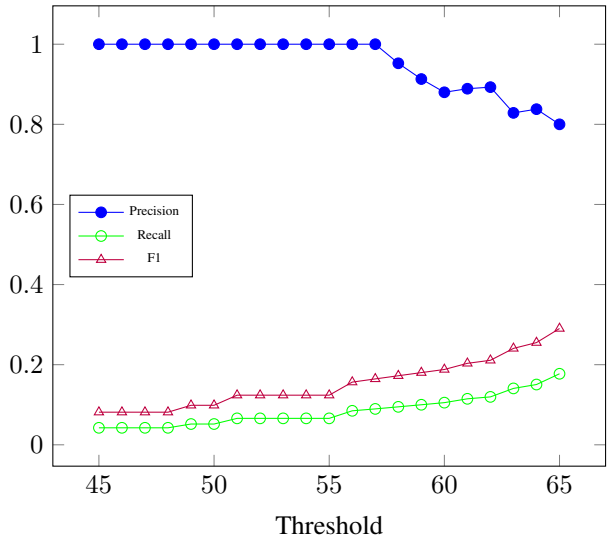

Fig. 3. Face recognition : influence of Binary Local Pattern parameter on precision and recall

is to offer correctly labeled data, we decided to trade recall for precision. For this first model, as depicted in Figure 6 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. We show in section V-B, how the continuous update of the model allows to overcome this low recall issue.

In FoP we demonstrate our approach on recognizing French politicians in national news. To boostrap our model, we therefore required an initial dataset with correctly labeled pictures. umerous celebrities are identified in the LOD-Cloud, mainly in DBPedia and Freebase. Regarding persons, both databases are linked together, not through the LOD-Cloud, but through the use of Wikipedia identifier.

For our needs, Freebase presented two main advantages. Firstly, the database has been manually generated and is therefore less error prone than the automatic generation of DBPedia. Secondly, for persons, Freebase usually provides links to several pictures which is not the case for DBPedia. We used the following approach to overcome potential issues in DBPedia.

We first extracted the list of living politicians in France along to their pictures from Freebase using the MQL query shown in figure 4. For each person, the number of pictures

[5]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.

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

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

[8]w.r.t. dbpedia updates

[9]*foaf* only

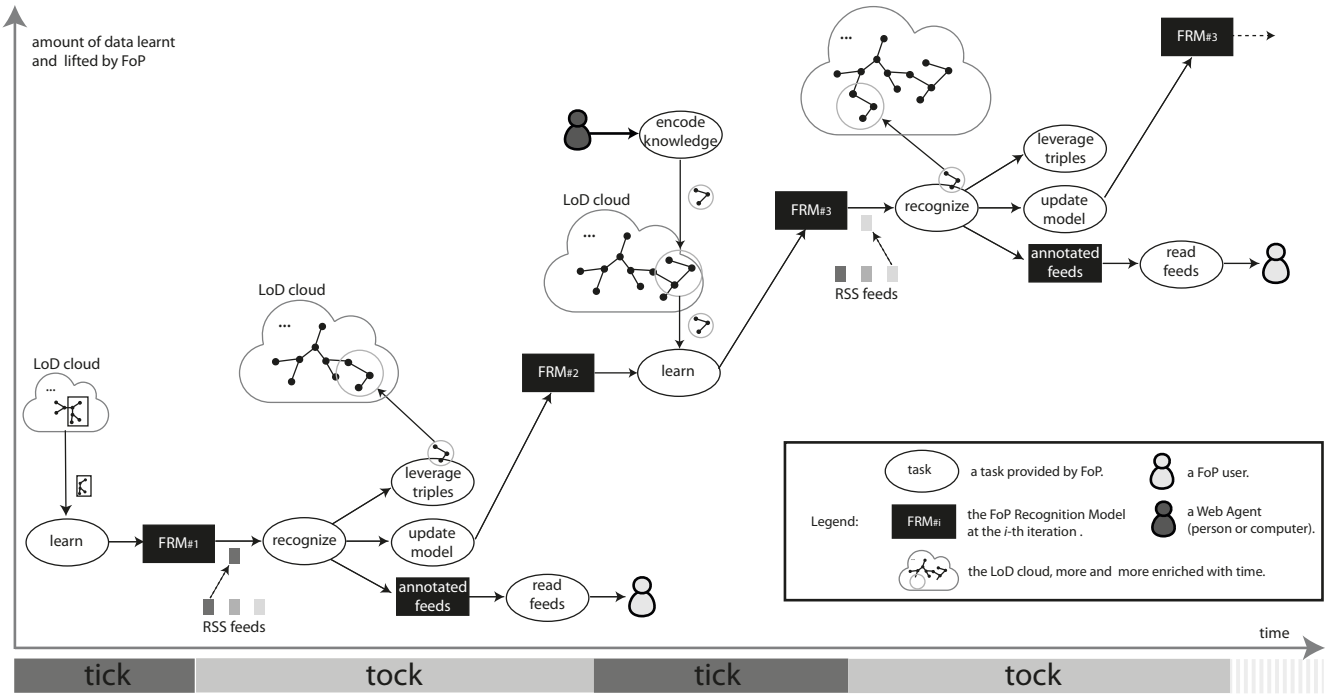

Fig. 2. Never-ending face recognizing and semantizing tick/tock model.

```
[{
  "mid": null,
  "name": null,
  "id": null,
  "type": "/government/politician",
  "/common/topic/image": [{
   "id": null
  }],
  "wiki_en:key": {
   "/type/key/namespace":"/wikipedia/en_id",
   "value": null
  },
  "ns0:type": {
   "id": "/people/deceased_person",
   "optional": "forbidden"
  },
  "/people/person/nationality": "France"
}]
```

Fig. 4. MQL Query to retrieve living french politicians along to their pictures and Wikipedia ID from Freebase

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. Then we retrieved for each person a second set of pictures from DBPedia using the Wikipedia identifier as previously stated. We used the previously trained model to assess the presence of the person and if validated, we updated the previous model.

V. Never-ending learning (steady state)

The never-ending learning part of FoP is made of two successive steps. In the first one FoP is querying the LoD-Cloud for new data and in the second one, it is providing new data to this latter. 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". The FoP system data leveraging flow is illustrated in Figure 2. 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.

A. Tocks

Tocks are data queried from the LoD-cloud that enrich the FoP learner. A first tock occurred when we initialized FoPas described in the previous section. However data from the LOD-Cloud are not static and are subject to change over time. For instance, new politicians may be added to Freebase or DBPedia, or new pictures may be added for a person. That is the reason why other tocks occur at regular interval of time. At further tocks, FoP queries against Freebase and DBPedia for new Picture–Person links, following the same process as described previously for newly added persons. For each of the new pictures that were introduced in Freebase or DBPedia since the last query, FoP updates the learnt model. FoP then updates its internal triple store 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. Conversely, 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

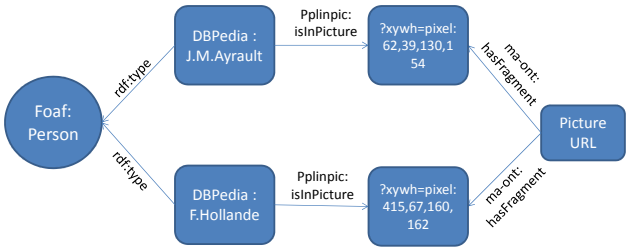

Fig. 5. People In Picture Ontology in use, for a picture having two people depicted. See Figure 8 for the SPARQL Query and Figure 9 for the resulting picture.

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 6, 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.

C. Providing linked data

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.

An depiction of the Ontology in use is given in Figure 5.

[10]http://www.w3.org/TR/mediaont-10/
[11]http://www.w3.org/TR/media-frags/

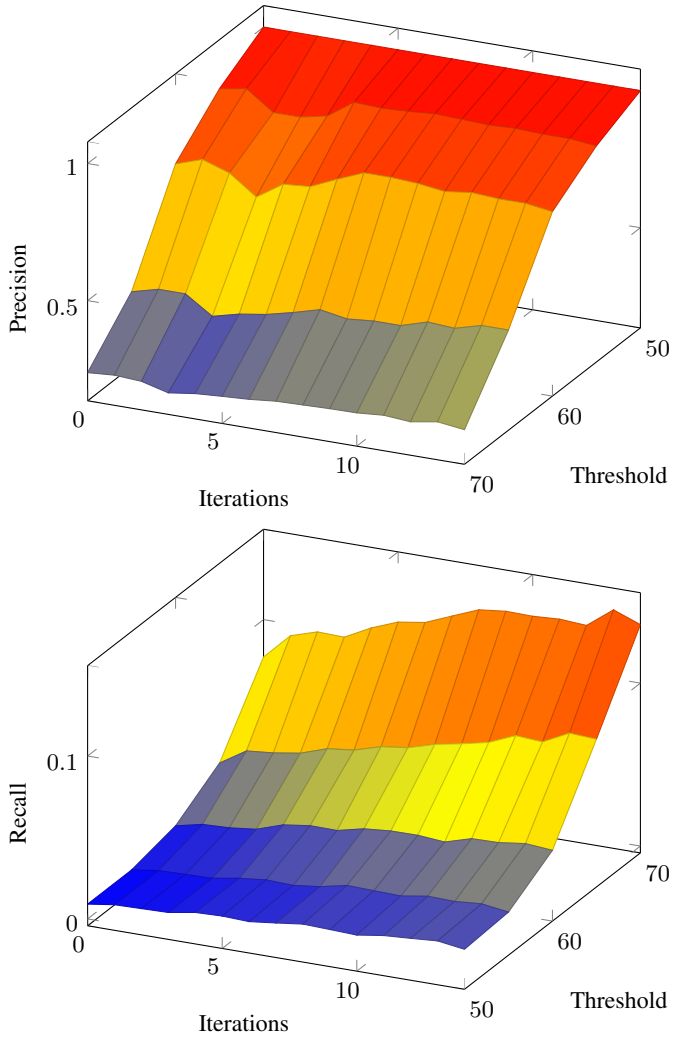

Fig. 6. Influence of updates on precision and recall. Note that for readibility, scale of threshold has been inverted between the two pictures.

This enables, among other, to search for pictures containing multiple people, as we demonstrate in the next section.

VI. Using FoP

Our FoP never ending learner is available online at http://demo-satin.univ-st-etienne.fr/facesofpolitics/. Figure 7 shows the current trained face recognition model fueling FoP. FoP is currently 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, plus the fine-grained representation that is then made possible.

On FoP Website, it is possible to browse latest annotated feeds. It offers a list of discovered persons along the part of the multimedia in which they were found, and the context of the original feed article. This is the primary task for a new FoP user in order to get familiar with the system.

A more advanced task is to make article curations on single or multiple person presence selection. By switching to

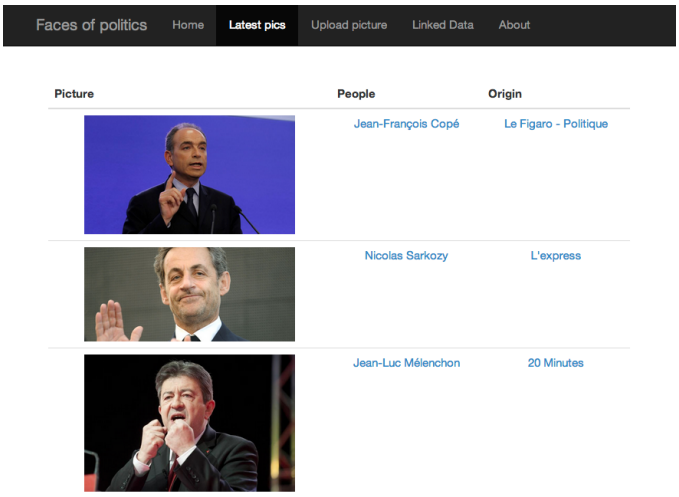

Fig. 7. Primary user interface for reading enriched RSS articles from subscribed feeds.

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 illustration of the article. Let us illustrate an example of a fine-grained query 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, French prime minister at the time of writing, is depicted left to François Hollande, French President, in all pictures in which FoP recognized both of them. This can be achieved using the SPARQL query[12] provided at Figure 8. The first article matching this query is given at Figure 9 as taken from one major French media website[13].

Full RDF Dumps are also available from the previously specified URL[14].

VII. Conclusion

In this paper we presented FoP, a never-ending face recognition learner that leverages annotated faces from RSS feeds through a continuous virtuous circle between the LoD-cloud and FoP recognition model.

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.

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. Finally, we provide a publicly available dataset of politicians pictures that is constantly growing and being enhanced.

In the future, we plan to study the evolution of FoP over time, especially the evolution of its precision and recall as it scales for more faces.We will aim at generalizing our system to foreign politician but also other applications like sport as well and larger categories of persons.

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

Fig. 9. Media retrieved by FoP using query provided by Figure 8

[12]The following prefixes were ommitted for formating considerations: 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#>

[13]http://www.europe1.fr/Politique/Chomage-la-methode-Coue-de-Hollande-et- Ayrault-1686655/

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

REFERENCES

[1] C. Matuszek, J. Cabral, M. J. Witbrock, and J. DeOliveira, "An introduction to the syntax and content of cyc." in *AAAI Spring Symposium: Formalizing and Compiling Background Knowledge and Its Applications to Knowledge Representation and Question Answering*. Citeseer, 2006, pp. 44–49.

[2] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor, "Freebase: a collaboratively created graph database for structuring human knowledge," in *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*. ACM, 2008, pp. 1247–1250.

[3] O. Nov, "What motivates wikipedians?" *Communications of the ACM*, vol. 50, no. 11, pp. 60–64, 2007.

[4] F. Wu and D. S. Weld, "Autonomously semantifying wikipedia," in *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*. ACM, 2007, pp. 41–50.

[5] D. Milne and I. H. Witten, "An open-source toolkit for mining wikipedia," *Artificial Intelligence*, vol. 194, pp. 222–239, 2013.

[6] M. Strube and S. P. Ponzetto, "Wikirelate! computing semantic relatedness using wikipedia," in *AAAI*, vol. 6, 2006, pp. 1419–1424.

[7] 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.

[8] A. Zimmermann, C. Gravier, J. Subercaze, and Q. Cruzille, "Nell2rdf: Read the web, and turn it into rdf," in *2nd Int Workshop on Knowledge Discovery and Data Mining Meets Linked Open Data, 10th ESWC 2013*, 2013, pp. 2–8.

[9] 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.

[10] 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.

[11] R. Troncy, E. Mannes, S. Pfeiffer, and D. Van Deursen, "Media fragments uri 1.0," W3C Working Draft 30 November 2011, W3C, Tech. Rep., November 2011. [Online]. Available: http://www.w3.org/2008/WebVideo/Fragments/WD-media-fragments-spec/

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

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

[14] 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.

[15] 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

[16] 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.

[17] 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.

[18] 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).

[19] 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/

[20] 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

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] A. F. Abate, M. Nappi, D. Riccio, and G. Sabatino, "2d and 3d face recognition: A survey," *Pattern Recognition Letters*, vol. 28, no. 14, pp. 1885–1906, 2007.

[26] X. Zou, J. Kittler, and K. Messer, "Illumination invariant face recognition: A survey," in *Biometrics: Theory, Applications, and Systems, 2007. BTAS 2007. First IEEE International Conference on*. IEEE, 2007, pp. 1–8.

[27] R. Jafri and H. R. Arabnia, "A survey of face recognition techniques." *JIPS*, vol. 5, no. 2, pp. 41–68, 2009.

[28] 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.

[29] 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.