SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Architecture of A Semantic Annotation of Handwritten Documents Based on the Ontology "OMOS"

[1]Mint Aboubekrine Aiche, [2]Mohamed Lamine Diakité, [3]Kamal Eddine EL Kadiri, [4]Youness Tabii
[1,3,4]LIROSA Laboratory; Systems Research Uni, ENSA , Tetouan, Abdelmalek Essaadi University, Morocco
[2]URDNI Research Unit, Faculty of Science and Technology, University of Nouakchott Alaasriya
ne3ma.20@gmail.com, diakite_medlamine@yahoo.com, elkadiri@uae.ma, youness.tabii@gmail.com

ABSTRACT

In this work we propose to present an application that supports the representation of manuscript documents according to an ontological approach. The implementation of this application makes it possible to annotate semantically these manuscripts according to the ontology "OMOS" [1]. The semantic annotation proposed here is an annotation based on an ontology of the "OMOS" manuscripts of generation and use of specific metadata targeted to allow new methods of access to information and to extend existing ones. The proposed annotation is based on the understanding that the named entities (the author, date 'link ... etc.) mentioned in the documents constitute an important part of their semantics. Finally, this application gives us an annotation file associated with each document.

Keywords: Semantic annotation, Ancient Manuscript, Ontology.

1 Introduction

As part of the safeguarding of the world's cultural heritage, digitization campaigns of ancient manuscript documents that constitute the cultural heritage of nations are becoming increasingly common. This digitization campaign generated a large number of digitized resources with an invaluable amount of knowledge. Likewise, a large number of ancient works and documents from all over the world are kept in archives which are threatened with disappearance due to moisture, acidity of paper, etc. It is therefore important to preserve this heritage, to make it accessible to everyone and especially with easy access. Digitization is the solution adopted, but it only provides images of documents, which is not always sufficient. Indeed, it is often necessary to access the contents of digitized documents and to exploit them possibly. This is the object of semantic annotation of these ancient manuscripts. To ensure the efficiency and reusability of metadata, we base ourselves on the OMOS ontology [1], it is based on specific knowledge of African handwritten documents, rather than being indifferent to any ontological commitment and any knowledge, but a vast knowledge base of the descriptive entities is maintained on the Arabic manuscripts. To carry out this work there are several state-of-the-art documents dealing with particular documents or documents of general application [3]. Here we present the Semantic annotation technique proposed. Finally, in this work we will describe the design of (ASDM) realized for the modeling of the semantic annotation based on the OMOS ontology. Our vision is that fully semi-automated methods for semantic annotation should be sought and developed. To do this, the necessary design and management issues must be extracted and annotated, and the necessary additional resources and

DOI: 10.14738/tmlai.54.3221
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3221

infrastructure must be provided. To ensure wide acceptance and proper use of semantic annotation systems their tasks must be clearly defined and their performance properly assessed and communicated.

2 Related Work

Several works on the semantic annotation exist in the literature. On the research side, several works exist for the semantic annotation among these works we mention some of them according to the field of research [3, 4, 5]. In another part of the research, a few works on ancient Arabic manuscripts exist [1], [6].First in the semantic web, semantic annotation is defined as the process that determines the interpretation of a document by associating it with formal and explicit semantics. This interpretation is commonly expressed in ontological terms when it is a question of associating a semantic type with the names of the entities mentioned in the text [7,8].The image format in which these documents are saved makes it difficult to be able to exploit their content satisfactorily, whereas current semantic annotation work exploits information retrieval techniques to associate document strings with of concepts of ontology [4].This annotation can be manual, automatic or semi-automatictexts. In return, we do not find an interested approach to the annotation of handwritten documents. The idea is, once digitized manuscripts, to facilitate their access to researchers and other users through a system and a thematic organization. As this annotation is expressed in ontological terms, we find that [1] OMOS ontology is the first one for Arabic heritage manuscripts. The objective is to find a way that will allow the needs of the users to be better interpreted and thus allow a better access to the content. To do this, we proposed a modeling of knowledge on manuscripts through the ontology "OMOS". Its objective is to implement an application for the annotation of ancient manuscripts. This application simplifies and facilitates access to these documents. Indeed, it exploits the descriptions associated with the image to construct semantic annotations

3 Document Semantic Annotation Technique

3.1 Description of a Document

Contrary to the classical representation of a document that only takes into account the textual content [3] and where the information is presented in a flat manner without any treatment, the structural presentation of a manuscript must enable us to identify At least two different levels of knowledge that characterize manuscripts.

The first level corresponds to the descriptive and situational knowledge that applies to the exploitation of the manuscript whatever its content.

The second level is linked to knowledge coming directly from the contents of the manuscript. Obviously, this type of knowledge is at the same time more difficult to obtain, [4] much more diversified and less limited than the previous level of description.

We started with the first level.

The following figure (1) presents the descriptive knowledge that could be used to properly present the manuscripts. This descriptive knowledge is available in the catalog of the IMRS library enriched by the suggestions obtained by the experts.

Information		About Resource	Type
Author	المؤلف	Manuscript	Concept
Title	العنوان	Manuscript/Copy/Copy with comments	Attribute
Subject	الموضوع	Manuscript/Copy/Copy with comments	Concept
Copyist	الناسخ	Copy	Concept
Write mode	الخطّ	Manuscript/Copy/Copy with comments	Attribute
Ink color	الحبر	Manuscript/Copy/Copy with comments	Attribute
Page surface	قياس الصفحة	Manuscript/Copy/Copy with comments	Attribute
Page number	الصفحات	Manuscript/Copy/Copy with comments	Attribute
Written surface	مساحة النص	Manuscript/Copy/Copy with comments	Attribute
Line number	الأسطر	Manuscript/Copy/Copy with comments	Attribute
Entire or not	النص تام ؟	Manuscript/Copy/Copy with comments	Attribute
Year[16] of publication	تاريخ النشر	Manuscript/Copy/Copy with comments	Attribute
Incipit[17]	البداية	Manuscript/Copy/Copy with comments	Attribute
Explicit[18]	النهاية	Manuscript/Copy/Copy with comments	Attribute
Exegetes		Copy with comments	Concept
Birth year / Death year		Author/Copyist/ Exegetes	Attribute
Birth Place		Author/Copyist/ Exegetes	Concept
Place of origin		Author/Copyist/ Exegetes	
Library		Manuscript/Copy/ Copy with comments	Concept
Location		Library / Private Library / Public Library	Concept
Language[19]		Manuscript/ Copy/Copy with comments	Concept
Field		Manuscript/ Copy/Copy with comments	Concept
Named time period		Manuscript/ Copy/Copy with comments	Concept

Figure 1: descriptive knowledge of manuscripts

3.2 Presentation of OMOS ontology

Ontologies traditionally play an important role in data integration processes [9],[10], which is also a key property when dealing with semantically heterogeneous sources such as handwritten documents. We present OMOS, an ontology describing West-Saharan manuscripts in Africa. The first version of an ontology on the manuscripts was realized "Semi-automatic construction of an ontology on Western Saharan manuscripts" we will try to use this ontology to semantically annotate the manuscripts efficiently. Our approach is based on the use of an OWL ontology that deals with the semantic web domain. This ontology is called OMOS, it is developed by the LABORATORY of INFORMATIQUE LI of the University François Rabelais of Tours in FRANCE. Is ontology contains hierarchies of concepts focusing on the domain of the semantic Web. Figure 2 presents the hierarchy of this ontology.

Figure 2: presents a hierarchical part of this ontology

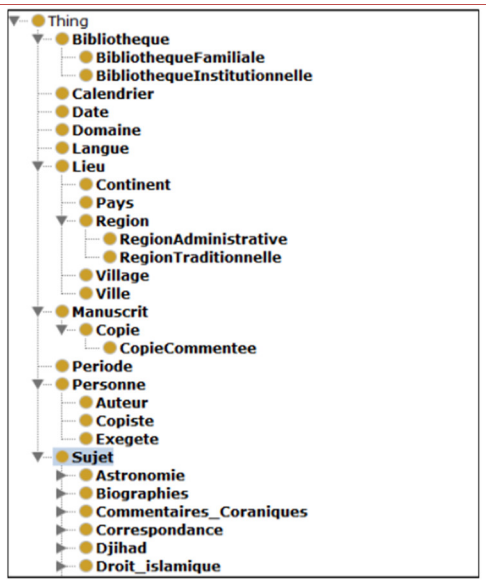

Figure 3: presents some of this ontology

3.3 Semantic annotation proposed

The semantic annotation of web resources is a way of moving from the current web to the vision of the future Web envisaged by Tim Berners-Lee. Indeed, it represents a process which aims to formalize semantic interpretations extracted from the document. The proposed technical process is a design of an application, which takes as an input step another approach. This application is a process that allows us to realize our object in order to offer us an annotated document. The latter determines a new document associated with semantic data.

4 Modeling Of A Sementicannotation

In this section we will present the architecture of our application and the implementation work we did, which consisted firstly of the edition of our ontology under the OWL language, followed by its exploitation in a semantic annotation application, which was developed in Java.

4.1 Semantic annotation architecture

In order to implement this application we propose a design presenting the structure of our work. In all cases, the semi-automatic annotation method will be adopted. The application will request the handwritten document of image format plus the information provided on this document andby monitoring the application can enrich these descriptions based on its own knowledge modeled in the form of the ontology of the domain. This will be discussed in the next section (see Figure 4).

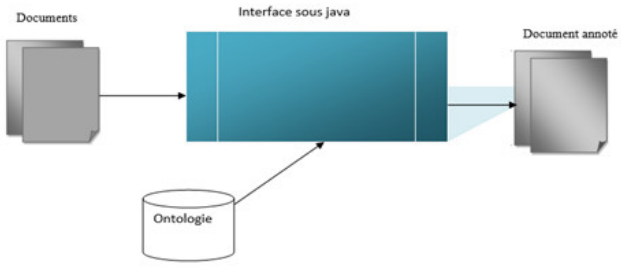

Figure 4: The architecture proposed for the application

4.2 Implementation

The implementation of our approach proposes an application implemented in JAVA and using the Jena library to exploit the different relationships. This JENA library contains a Java API for the manipulation of RDF / RDFS and OWL files, for the diffusion of ontology and the management of queries.

5 Result

These methods mainly use two concepts, namely the ontological approach to semantically apply and model the metadata, and the annotation application that allows to associate interpretations with the handwritten documents and facilitates the communication between its different Components. Another interesting line of research to be devoted to it is the identification and the extraction of the information of the manuscript like the name of the author, the copies and date of creation and the place etc. This information can then be semantically annotated according to an OMOS ontology-based application. This allows you to search annotated information.

6 Conclusion

We chose this methodology as it allowed to formalize and generate all the explicit and implicit knowledge on the manuscript and to annotate it semantically according to the OMOS ontology.

As research perspectives, we have identified this theme that it would be interesting to explore. The proposed application may ultimately serve as a basis for research in manuscripts.

The different concepts defined there will serve as a semantic repository to better interpret user queries.

REFERENCES

[1] Mohamed Lamine DiakitéBéatrice Bouchou Markhoff, « Construction semi-automatique d'une ontologie sur des manuscrits ouest sahariens » Communication dans un congrès IC2015, Jun 2015, Rennes, France. AFIA.

[2] Yue Ma, Laurent Audibert, Adeline Nazarenko « Ontologies étendues pour l'annotation sémantique » 20es Journ_ees Francophones d'Ing_enierie des Connaissances, May 2009, Hammamet, Tunisie, Tunisie. pp.205-216, 2009. <hal-00378594>

[3] A. Guissé, F. Lévy, A. Nazarenko, and S. Szulman «Annotation sémantique pour l'indexation de règles métiers » Conférence Internationale sur la Terminologie et l'Intelligence Artificielle (TIA 2009), page (electronic medium). Université Paul Sabatier - Toulouse, (November 2009).

[4] UREN V., CIMIANO P., IRIA J., HANDSCHUH S., VARGAS-VERA M., MOTTA E. & CIRAVEGNA F. (2006). Semantic annotation for knowledge management.

[5] Teyeb, M. Torjmen, N. Hernandez, O. Haemmerlé, and M. Jemaa. « Vers une annotation sémantique des images web fondée sur des patrons RDF ». CORIA, page 285 300. ARIA, (2015).

[6] Bertrand Coüasnon, Jean Camillerapp, « Accès par le contenu aux documents manuscrits d'archives numérisés », Document numérique 2003/3 (Vol. 7), p. 61-84 DOI 10.3166/dn.7.3-4.61-84.

[7] KIRYAKOV A., POPOV B., TERZIEV I., MANOV D. & OGNYANOFF D. (2004). Semantic annotation, indexing, and retrieval. J. Web Sem., 2(1), 49–79.

[8] Nguyen M., « Vers une plate-forme d'annotations sémantiques automatiques à partir de documents multimédias », 2007.

[9] H.Wache, T.Vögele, U.Visser, H.Stuckenschmidt, G.Schuster, H.Neumann, and S.Hübner, Ontology- Based Integration of Information – A Survey of Existing Approaches. IJCAI Workshop on Ontologies and Informations Sharing, pp 108–117, 2001.

[10] Atanas Kiryakov, Borislav Popov, Ivan Terziev, Dimitar Manov, Damyan Ognyanoff « Semantic annotation, indexing, and retrieval » Web Semantics: Science, Services and Agents on the World Wide Web, Volume 2, Issue 1, Pages 49-79.