International Journal of Information Systems in the Service Sector
Volume 8 • Issue 2 • April-June 2016

Web Application for User Profiling

Taous Iggui, LIMED Laboratory, Department of Computer Science, University of A. Mira, Bejaia, Algeria

Hassina Nacer, Department of Computer Science, University of Science and Technology Houari Boumediene (USTHB) Algiers, Algeria

Youcef Sklab, LIMED Laboratory, Departement of Computer Science, University of A. Mira, Bejaia, Algeria

Taklit Ait Radi, Department of Computer Science, University of A. Mira, Bejaia, Algeria

ABSTRACT

User's profiles play an important role when information systems try to meet their needs. This work presents a novel approach to build user profiles. It is based on information extraction techniques and proceeds by iterative steps. The use of different statistic metrics, Natural Language Processing (NLP) techniques and semantic descriptions (ontologies) in the authors' approach, has provided it with a good precision degree when extracting information from texts. This has been demonstrated by an application prototype which is an automatic user profile constructor, using the texts of emails job applications (E recruitment field).

KEYWORDS

Information Extraction, Natural Language Processing, Ontologies, Semantic Web, Statistics, User Profile

1. INTRODUCTION

Information systems are, increasingly, accessible via Internet or Intranet, they allow users to reach an enormous mass of information from several and different sources. The most important challenge in current researches deals on how to optimize the performance and the accuracy of the results returned by applications and computer systems, in order to better satisfy the user's needs (Ryan & Finn, 2005; Weib et al, 2008).In order to meet user's needs (Harrathi, 2009), the most important challenge in current researches deals on how to optimize the performance and the accuracy of the results returned by applications and computer systems. The integration and consideration of the semantic aspect in information processing is extremely important to better understand user's requests. To do this, a set of constraints is considered, and *user's profile* (Schiaffino & Amandi, 2009) is a decisive one (Lee, 2009). It is a set of information concerning the user (personal data, education, professional information, preferences, interests …etc.).

The user profile is a crucial concept for various fields, such: e-recruitment, e-commerce, criminal domain, information retrieval, education, e-learning, security (Kaczmarek et al, 2010; Jeong & Choi, 2012; Satler et al, 2010) and many others. It's related to several disciplines such as: Artificial Intelligence (AI), Language Processing, Text-mining and Semantic Web.

In this context and given the great impact of using the user profile by several systems and applications, the authors, in this paper, propose an information extraction system to construct *user's profile*, from electronic e-mails of job applications (in the e-recruitment domain), by combining between several statistical metrics, inspired by works of (Bossard, 2010; Boudin & Torres, 2009) semantic Web technologies, as well as NLP (Natural language Processing) techniques and methods.

DOI: 10.4018/IJISSS.2016040104

So, the system can take into account all the relevant information about user's preferences and activities. The remainder of the document is organized as follows:

In section 2, the authors synthesize the most important works in the field of Information Retrieval (IR), Information Extraction (IE) and building a user profile. In Section 3, they propose their system architecture, entitled PBS (Profile Builder System), an automatic system to construct user's profile and they detail the various component modules. Section 4 illustrates the evaluation of the proposed approach, by presenting results and performance measures of this one. Finally, conclusion and future works are described in Section 5.

2. RELATED WORKS

Information Extraction (IE) aims at exploring and exploiting various formats of data. It uses a set of techniques and methods to extract relevant information that can be conveyed by an information source. Several research studies have emerged, in this section we present a classification of the most relevant works according to their application field:

2.1. Automatic Summarization

An automatic summarization system aims at returning a condensed representation of an original text while keeping its semantic. In other words, it aims at extracting relevant information. Several summarization techniques have been developed and allow the construction of several summarization systems such (Torres & Rodrigez, 2010; Bossard, 2010; Boudin, 2008; Erkan & Radev, 2004), YACHS2, CORTEX and ROUGE2.

They use a set graph theory techniques and statistical metrics. Once assembled, they allow to assign a score to each sentence of the document, then, sentences having the highest scores are selected to be a part of the summary.

2.2. Information Retrieval

The arising problem in the field of information retrieval is about how to get relevant information to meet user's needs according his preferences and interests.

Authors like (Sieg et al; 2007) have used user profiles for a personalized information search. A user profile is used in the results ranking step. In this area, generally, the user profile is represented by a set of key words, extracted from consulted documents by using the tf.idf measure (Tamine et al, 2007), ontologies (Daoud et al, 2008), and other knowledge bases like Wikipedia in (Nicoletti et al, 2012).

2.3. Information Extraction (IE)

Actually the intersection of methods and techniques, offered by IE discipline and the user profile concept, gives rise to the topic of building this user profile. Techniques of IE allow us to construct the user profile, from a set of relevant key words extracted from applications titles (Seguela et al, 2010), using a set of named entities (Omrane et al, 2011), from its comments in Web forums (El manar el bouanani & Kassou, 2012) by using a set of clustering techniques (K-means, EM), methods of stylometry as well as ontologies of speech.

In the recruitment area, (Kaczmarek et al, 2010) treats documents in Polish language, by exploiting extraction rules and an XML representation.

2.4. Reviews and Discussion

During the study and the analysis of the previous works, the authors have been able to identify the characteristics of each method. They noticed that the extraction of relevant information either the construction of a user profile is based itself on a set of statistical metrics, which provide the highly-rated formal of the obtained results. But they also underline the independence between these results and treated texts, in other words ”the absence of semantic constraint”. In fact, the statistical metrics return approximate results, but require a semantic validation, for example, a too frequent term may not be relevant.

There is also, works treating the constraint of the evolution of the user profile, based on his interactions with the system, which still don’t reflect the user’s interests.

A user profile is more important than a set of key words and user interactions; it’s a set of semantic dimensions.

3. PROPOSED APPROACH: PROFILE BUILDER SYSTEM (*PBS*)

For the construction of a user profile, we have turned to an area where the user profile has a great impact, which is the area of e-recruitment. Certainly, this one allows solving the problem of overload recruiters and administrators with traditional methods, but the complexity of the treatment is quite huge.

Having for purpose, saving time and improving the e-recruitment process quality, the authors propose an automatic system for the construction of a user profile from emails’ texts of job applications.

Figure 1 shows the proposed system architecture.

The user profile will be represented according to the user profile model of (Sklab et al, 2013), illustrated in Figure 2.

Figure 1. The proposed system architecture

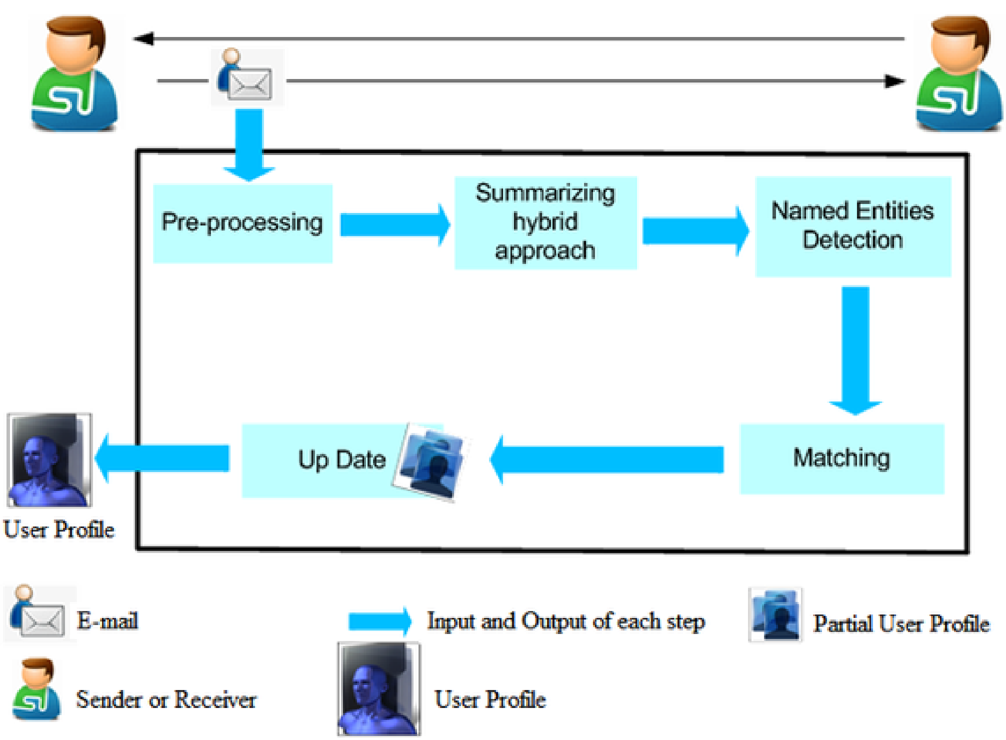
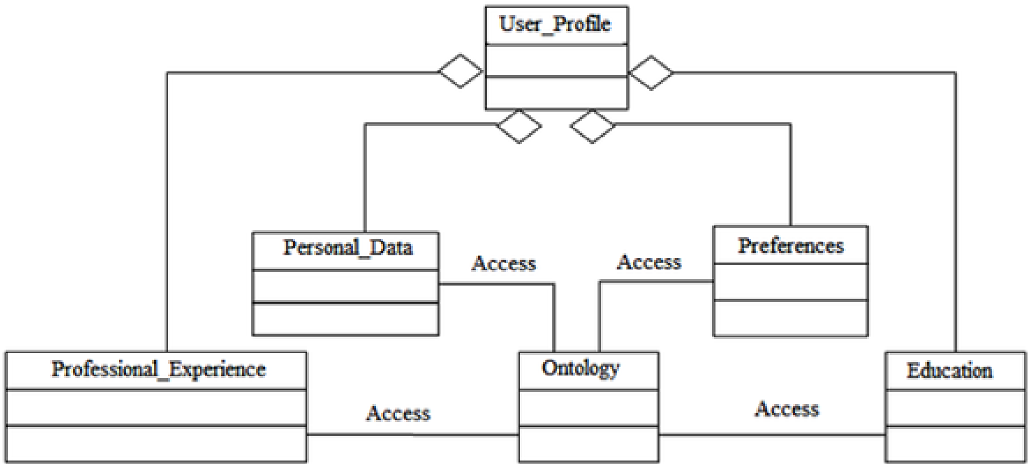


Figure 2. The user profile model



The model is composed of a set of semantic dimensions guided by an ontology.

The different modules of this system are as follows.

3.1. Presprocessing

So as to facilitate handling and managing textual information, the researchers use XML technology, due to different benefits offered by this representation (Min et al, 2003): A structured and organized model, easy to implement, an excellent format for exchanging Data, and possibility of a relational data representation.

The treatment involves a set of operations such: segmentation, normalization and terms lemmatization.

3.2. Hybrid Summary

In this module, the authors propose a summarization method *method_A*, which is a combination of a set of statistical metrics, having already proved their worth in the automatic summarization field, with domain ontologies. This combination ensures a summary oriented by the ontology. The *method_A* will be coupled with *method_B* of (Torres & Rodrigez, 2010) based on graphs and social networks prestige.

The output of the hybrid summary module will be the union of the results of *method_A* and *method_B* (Torres & Rodrigez, 2010).

Adding a summary module to treat the user's emails and extract the user profile information is an original idea. The goal, here, is to save time and avoid wrong *attributes assignment (this will be discussed in the next modules)*.The relevant sentences returned by the previous module represent the set of the most probable sentences to convey information concerning the user profile.

3.3. Named Entities Determination

Once the most relevant sentences are selected, the authors draw on (Bouhaf & Desclé, 2005), to determine the Named Entities (NE) convoyed par the email. The goal of this step is to locate and classify text elements into predefined categories corresponding to the nature of the user profile attributes. To do this, they use a set of Named Entities Detection Rules (NEDR).

3.4. Matching with User's Profile Information

This step aims at associating each detected NE to the corresponding user profile attribute. The matching operation is done in three phases:

3.4.1. Determination of Information's Unit

Each attribute of the user profile model is related to a specific topic with the intermediary of a triggering element. Based on thisidea and inspired by (Genest & Lapalme, 2011), the authorsuse analgorithm ofgrammatical dependencies between words. These dependenciesare represented by predicates, PREDICATE(Word1, Word2), where Word1, Word2 belong to the sentence and PREDICATE indicates the nature of the relation between these two words. This predicate has true value, if the relation between the two words exists, else false. At the end of this step, the result will be a set of basic units, that can provide relevant information about the user profile, called a set of "*Information Unit*" (IU).

3.4.2. Selection of Candidate Attributes

The trigger verb may be common to several attributes, or in other words, several Named Entities belonging to different categories may be identified by the same trigger verb. For example, the verb to be in "I am Johnny" introduces the *name,* while in "I am 25 years old" it introduces the *age* of the person. The question here is: how to explain this semantic to a machine(to the system) ? The aim of this step is to select the most appropriate attribute to receive the information or the *attribute_value* conveyed by the Information Unit. To do this, the researchersdefine the similarity measure between attributes and the IU.

3.4.3. Validation

The *attribute_value* conveyed by anInformation Unit may have severalcandidate attributes, from the user profile model. Hence, the need for the validation step, which consists into the validation of the right attribute. To do this, the researchers propose a set of generic rules to decide about the information categories of the user profile model, called *"Contextual Exploration Rules"*.

Application Example:

To better illustrate the validation step, the authors present the following example, having the sentence: *"I live in London city and I work with IBM"*.

The system can distinguish two information units:

- I live in London city.
- I work with IBM.

Taking as an example the first one, the validation will be as follows:

City belongs to the semantic indicators (IND) of the attribute *residence* from the user profile model, so *London* will be allocated to it.

3.5. Updating the User Profile

Considering the evolution characteristic of the user profile, the authors suggest the updating module. It aims to update the information of the user profile, while respecting the scalability constraint according to the nature of each attribute (*static attributes, dynamic attributes and scalable attributes*).

We define:

StaticInfo: Represents a set of static attributes of the user profile.
DynamicInfo: Represents a set of dynamic attributes of the user profile.
ScalableInfo: Represents a set of scalable attributes of the user profile.
A: Represents a user profile attribute.

J: Means the j^{th} execution of the system.
Profile$_j$(Ai): Means the i^{th} attribute of the j^{th} execution.

3.5.1. Static Attributes

This kind of information (attributes) doesn't change, such as personal data (name, birthday and nationality).

$$\forall j, \forall Ai \in StaticInfo, \Pr ofilej\left(Ai\right) == \Pr ofilej+1\left(Ai\right)$$

3.5.2. Dynamic Attributes

In this case, the researchers associate to each dynamic attribute the most current value detected, in other words, the i^{th} attribute of the j^{th} execution will receive the returned value by the $(j+1)^{th}$ execution, such as: age and family situation.

$$\forall j, \forall Ai \in DynamicInfo, \Pr ofilej\left(Ai\right) \leftarrow \Pr ofilej+1\left(Ai\right)$$

3.5.3. Scalable Attributes

Here, the attribute values are obtained by the union of the values returned by each partial profile, in other words, it's the union of the different values returned by all the previous executions. Some of the scalable attributes of the constructed user profile: experience, function, diploma…etc.

$$\forall j, \forall Ai \in ScalableInfo, \Pr ofilej+1\left(Ai\right) \leftarrow \\ \Pr ofilej(Ai) \bigcup \Pr ofilej+1\left(Ai\right)$$

4. SYSTEM'S RESULTS

After presenting the different modules of the approach and their different details, the validation of the returned results is needed. In this section, the authors introduce some executions in order to evaluate the proposed system, supposing that:

- The emails are written in English.
- The email's body is not an attachment.
- The information shall be expressed by a single sentence.
- The sentences are not in the negative form.
  They also use an ontology in the computing recruitment field.

4.1. Scenario N° 01: Execution of the Approach Without the Module of "Hybrid Summary"

In this case, the goal is to identify the most relevant information conveyed by the email's sentences, according to the existence of IU.

After the execution of the third module of this approach, the application of a set of Named Entities Detection Rules (NEDR), on an example email, the researchers get the result presented in Figure 4, where each NE identified is annotated by the information category is belonging to.

Each detected named entity will be assigned to the corresponding attribute of the user profile model by the matching module.

Table 1 illustrates the application of the IU acquisition algorithm. The goal is to determine the set of information unit convoyed by an email. To do this, the system has to identify the different grammatical dependencies between terms, while based on Stanford[1] dependencies, which is based on the triggers verbs that guide the semantics' of the sentence. For example, having the predicate: nsubj (graduated, I) means that the verb *graduated* is related to the subject *I*.

The authors note: Rk *as the rank of the sentence.* V.Trig as *the trigger verb.* Dep.Subj *as subject dependencies and* Dep.Obj *as the* o*bject dependencies*.

The IU acquisition algorithm returns the set of Information Unit contained in each sentence.

A complete sentence is, generally, composed of a subject, a verb and an object, all linked by different grammatical dependencies. After the identification of the different information unit, each verb and subject will receive a set of objects, but only named entities NE are taken into account. These selected named entities represent relevant information that can be added to the user profile as an attribute value, but the assignment of a named entity requires the selection of the right attribute.

Table 2 shows that for each object or named entity (or attribute_value) there is a set of competing attributes, which are all the attributes having the same verb trigger and the same type of information of the IU object (NE).

Indeed, Table 2, represents a subset of pairs *(attribute, value)* detected in some sentences of the email, Figure 4. The selection of a candidate attribute is ensured by a similarity measure. For this, the researchers use *Word-Net::Similarity 2.05*, a tool, used to calculate the similarity between two terms using the WordNet knowledge base.

The researchers note: Rk *as the sentence's rank*. Ctx *the context of the Information Unit.*A.CC *the Competitors Attributes.* S.IU: *the similarity to the Information Unit*. A.C: *the Attribute Candidate and* A.W *the winner attribute* (after validation).

- - - *:* means that the attribute is not taken into consideration.

After, the selection of the attributes candidates, the authors noted some cases where a named entity had more than one attribute candidate, as it is illustrated in the first information unit of the second sentence of the email, in Table 2.

Having John Smith as a named entity, the attributes Name and JobTitle, having both the same degree of similarity0.16, they are both selected, by the system, as attributes candidates.

Semantically absurd, the authors propose the validation step, which consists into the validation of the right attribute, by the application of a set of generic rules that they have assembled according to the information's categories of the user profile, called "Contextual Exploration Rules" as it is shown in Figure 3.

After the validation step, with a set of contextual exploration rules, the system result is shown in Figure 5 as a partial profile, composed by the different values assigned by the previous module:

During the execution of this approach, the researchers have noticed and detected some special cases of attirubtion errors.

In Table 3, the authors present some examples (some sentences) in order to illustrate some special executions of this system, which are as follows:

1. I am Bill.
2. I have got a diploma from MIT.
3. I have done a training in MIT.

Let's note:*GA/BA* as Good Attribute/Bad Attribute, respectively.

Table 1. Acquisition of the information units

Rk	V. Trig	Dep. Subj	Dep. Obj
2	Am	Advmod (am, I)	Nsubj (am, John)
	Come	Partmod(I, come)	Prep_from (come, Algeria)
3	Born	Advmod (born, I)	Prep_on (born, April)
4	Gradua-ted	Nsubj (gradu-ated, I)	Prep_from (graduated, MIT) Prep_in (graduated, June)
	Have	Nsubj (have, I)	Nn (degree, Master) Dob (have, degree)

Table 2. Selection of the attributes candidates

Rk	Ctx	A.CC	S.IU	A.C	A.W
2	Am, Mr., John, Smith	Name	0.16	Yes	Name
		Nationality	0.11	--	
		Family-Situation	0.14	--	
		JobTitle	0.16	Yes	
		Degree	0.13	--	
		Domain	0.11	--	
		Name-Institution	0.12	--	
	Come, from, Algeria	Residence	--	Yes	Residence
3	Born, on, April	Birthday	--	Yes	Birthday
4	I, graduated, from, MIT, in	Name-Institution	0.15	Yes	Name-Institution
		Degree	0.12	--	
		Domain	0.10	--	
	I, graduated, from, in, June	Start-Date	0.07	--	EndDate
		EndDate	0.15	Yes	
	Have, a, master, degree	Degree	0.24	Yes	Degree
		Family-Situation	0.12	--	

Figure 3. The validation step

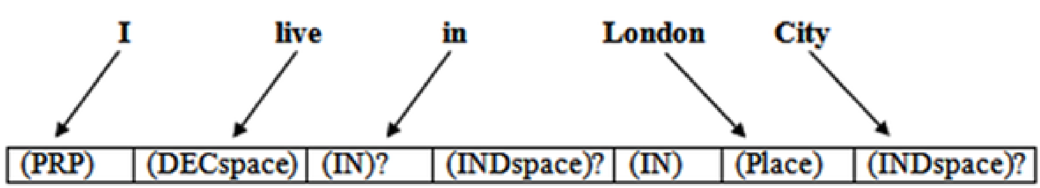


Figure 4. The detection of the named entities

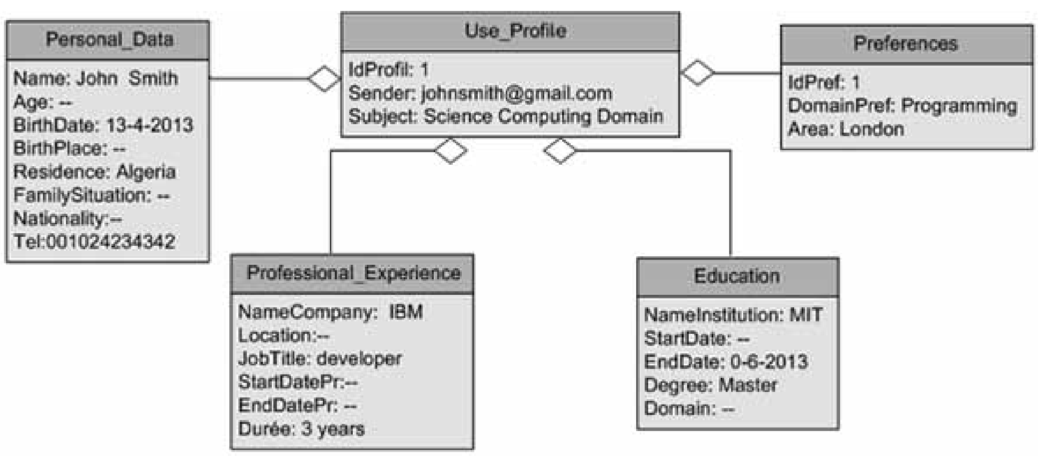
Dear, I am writing this email after reading your employment announce, in <Name> The Sun </Name> newspaper, in order to work with your company.

I am Mr. <Name><ProperName> John Smith </ProperName></Name>, I come <Prep> from </Prep> <Space><Countr> Algeria </Countr> </Space>. I was born <Prep> on </Prep> <Date> <Month> April </Month><NbDay> 13 </NbDay>, <NbYear> 1990 </NbYear> </Date>. I was graduated <Prep> from </Prep> <Name><Company> MIT </Company> </Name> <Prep> in </Prep> <Date><Month> June </Month> <NbYear> 2013 </NbYear></Date>, I have a <Name> <Degree> Master</Degree></Name> degree. Actually, I live <Prep>in</Prep> <Space> <Countr> London </Countr> </Space>, I have worked <Prep> with </Prep> <Name> <Company> IBM </Company> </Name> company <Prep>for </Prep><Period> <Numbr> three </Numbr> <NotnT> years </NotnT> </Period>, as a <Name> <Spetiality> developer </Spetiality> </Name>. I am interested on <Name> <Domain> Web </Name> </Domain>, <Name><Domain> Artificial Intelligence </Name> </Domain> and <Name> <Domain> Programming </Name> </Domain>. I can be contacted at <Num>001 024 234 342</Num>.

Thank you for your time and consideration, I look forward to speaking with you about this employment opportunity, sincerely.

Name Date Period Country

Figure 5. UML diagram of the returned user profile



The researchers note from Table 2 and Table 3, that the accuracy of the system depends, in large part, on the detection of the information unit and on the similarity measure. Indeed, they have distinguished some special cases, which may be or cause incorrect attributions. Some of them will be treated in the future works. And to the others, they have, already, proposed some solutions, as follows:

- **Case 01:** The attributes *StartDate* and *EndDate*, being of the same kind (Date) and having the same verbs triggers and the same semantic indicators, the similarity measure may be not significant (they can have the same degree of similarity).

Table 3. Other application cases

Rk	Ctx	A.CC	S.IU	A.W	GA/BA
1	I, am, Bill	Name	0.2491	JobTitle	BA
		Nationality	0.1687		
		Family- Situation	0.1655		
		JobTitle	0.2545		
2	Have, get, from, a, MIT	Name-Institution	0.1111	Degree	BA
		Domain	0.1012		
		Degree	0.1385		
3	Have, in, done, a, MIT	Institution	0.2045	Domain	BA
		Domain	0.2328		
		Company	0.1489		

- *Proposed solution: C*omparison of the two dates.
- **Case 02:***"I am Bill, I am Algeria, I am a Master degree",* the authors noticed in these examples, the semantic ambiguities between the information of the same type, as illustrated in the previous examples, Table 3, where the attribute Name undergoes an error attribution.
- *Proposed solution:* Use of a thesaurus, as a set of lists of attribute values (degree, domain, residence, company, nationality, jobTitle).

4.2. Scenario N° 02: Execution of the Approach With the Module of "Hybrid Summary"

Here, the researchers present the results returned by the execution of the module of "Hybrid Summary ".

The 8 sentences of the example email, of Figure 4, are numbered according to their appearance rank, as follows:

1. Dear, I am writing this email after reading your employment announces, in The Sun newspaper, in order to work with your great company.
2. I am Mr. John Smith, I come from Algeria.
3. I was born on April 13, 1990.
4. I was graduated from MIT in June, 2013, I have a Magister degree.
5. Actually, I live in London; I have worked with IBM company for three years, as a developer.
6. I am interested in the Web, Artificial Intelligence and programming.
7. I can be contacted at 001 024 234 342.
8. Thank you for your time and consideration, I look forward to speaking with you about this employment opportunity, sincerely.

The Hybrid Summary applied to the example email, Figure 4, gave the following results:

- *Method_A* = {2, 3, 4, 5, 6, 7}.
- *Method_B* = {1, 5, 4, 6, 2, 7}.
- *Hybrid Summary result* = {1, 2, 3, 4, 5, 6,7}.

The authors notice that the summary method, they have proposed *(Method_A)*,using domain ontologies, gave, indeed, a better result, by eliminating the sentences 1 and 8, which are sentences of greetings.

While with the *Method_B*, by absence of semantic, the sentence 3 has been eliminated in spite of its relevance, while sentence 1, which conveys no relevant information for the user profile, has been kept.

To avoid any loss of relevant information, the final result of the hybrid summary is the union of the results returned by *Method_A* and *Method_B*.

4.3. Evaluation Metrics of the Proposed System

Given the time and the complexity of the implementation of this approach, the researchers improvised a manual execution with five e-mails with very complex sentences, so allowing the validation of the expected results. The results of this set of emails, were quite satisfactory and very convincing, they take for reference, the quality of the results returned by the tools used in the implementation step, and the obtained values of the evaluation metrics. Among the metrics they were able to define:

- *Precision* = Number of attributes well detected /Number of attributes detected.
- *Error rate* = 1 - Precision.
- *Efficiency* = Number of attributes well detected (in an email) / Number of attributes conveyed by the email.
- The relation between the number of information unit and the accuracy of the system.

In Table 4 the researchers present the precision values, obtained during the different executions of their system, according to the number of sentences and information unit of each email:

5. CONCLUSION

The researchers propose in this work, an automatic system to construct a user profile from textual documents, called PBS (*Profile Builder System*), but also, a new domain-oriented automatic summary method, supporting semantics and minimizing the time of processing. The authors have executed the proposed approach on a set of electronic emails, taking the example of the applications in the IT domain. They have detected a convincing set of Named Entities, so allowing the construction of a partial user profile (by the Matching module), which become global after one or several updates. The returned results were rather satisfactory, concerning the results of every module and the approach generally.

During, the execution of this approach, the authors were able to distinguish a few scenarios that they envisage as perspectives, and among:

- The treatment of documents sent as an attachment.
- The construction of a user profile from a multi-domain email.
- The integration of a lexical analyzer based on semantics.

Table 4. The Proposed System Evaluation

	Sentences number	IU's number	Precision per E-mail	Error rate
E-mail 1	8	14	100%	0%
E-mail 2	4	5	40%	60%
E-mail 3	4	6	50%	50%
E-mail 4	4	6	66%	34%
E-mail 5	3	7	71%	29%
	The System accuracy		64.4%	

REFERENCES

Bossard, A. (2010). Contribution au résumé automatique multi-documents [Thèse de Doctorat]. l'Université Paris-13, Paris-Nord.

Boudin, F. (2008). Eploration d'approches statistiques pour le résumé automatique de texte [Thèse de Doctorat] l'Université d'Avignon et des Pays de Vaucluse. France.

Boudin, F., & Torres, M. J. M. (2009, Juin). Résumé automatique multi-document et indépendance de la langue: une première évaluation en français. Proceeding de la 9ieme Confrence annuelle sur le Traitement Automatique des Langues (TALN'09). Senlis, France.

Bouhafs, A., & Desclé, J. P. (2005). Utilisation de la méthode d'exploration contextuelle pour une extraction d'informations sur le web dédiée à la veille, réalisation du système informatique java veille [Thèse de Doctorat]. l'Université Paris IV Sorbonne, France.

Daoud, M., Tamine, L., Boughanem, M., & Chibaro, B. (2008). Construction des profils utilisateurs à base d'ontologie pour une recherche d'information personnalisée. Journal of Laboratoire IRIT Universit'e Paul Sabatier. France.

El manar el bouanani, S., & Kassou, I. (2012). Vers une méthodologie de Modélisation dune signature unique des profils web: Module de détection des auteurs des forums web [Thèse de Doctorat]. l'Université Mohammed V Souissi, Rabat.

Erkan, G., & Radev, D. (2004). Graph-based centrality as salience in text summarization. *Journal of Artificial Intelligence Research*, 22(1), 39.

Genest, P. E., & Lapalme, G. (2011, 24 June). Framework for abstractive summarization using text-to-text generation. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (Workshop)* (pp. 64-73). Portland, Oregon.

Harrathi, F. (2009, Septembre). Extraction de concepts et de relations entre concepts à partir des documents multilingues: Approche statistique et ontologique [Thèse de Doctorat] Nationale des Sciences Appliquées de Lyon. France.

Jeong, H., & Choi, E. (2012, Septembre). User Authentification using Profiling in Mobile Cloud Computing. Proceedings of the AASRI Conference on Power and Energy Systems, Hong Kong (pp. 262-267).

Kaczmarek, T., Zyskowski, D., Walczak, A., & Abramowicz, W. (2010). Information extraction from web pages for the needs of expert finding. Studies in Logic. *Grammar and Rhetoric*, *22*(35), 141–157.

Lee, D. L. (2009, November). User Profiling and Personalized Information Delivery on the Static and Mobile Web. *Proceedings of the eleventh international workshop on Web information and data management WIDM'09*, Hong Kong, China (pp. 1-2). ACM. doi:10.1145/1651587.1651589

Min, J. K., Ahn, J. Y., & Chung, C. W. (2003, January). Efficient extraction of schemas for xml documents. *Information Processing Letters*, *85*(1), 7–12. doi:10.1016/S0020-0190(02)00345-9

Nicoletti, M., Schiaffino, S., & Godoy, D. (2012). Mining interests for user profiling in electronic conversations. *Expert Systems with Applications*, *40*(2), 638–645. doi:10.1016/j.eswa.2012.07.075

Omrane, N., Nazarenko, A., & Szulman, S. (2011, Septembre). Le poids des entités nommées dans le filtrage des termes d'un domaine. Proceedings of 9th internationale de Terminologie et Intelligence Artificielle, Paris, France.

Ryan, P., & Finn, E. (2005). Field-based mLearning: who wants what? Proceedings of the 7th international conference on Human computer interaction with mobile devices & services MobileHCI'05 (pp. 327-328). ACM.

Satler, M. F., Romero, F. P., Menendez, V. H., Zapata, A., & Prieto, M. E. (2010, July). A Fuzzy Ontology Approach to represent User Profiles in E-Learning Environments. *Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ)*, Barcelona.

Schiaffino, S., & Amandi, A. (2009). Intelligent User Profiling. Artificial Intelligence, LNAI (Vol. 5640, pp. 193 – 216). Berlin. doi:10.1007/978-3-642-03226-4_11

Séguela, J., Saporta, G., & Stéphane, L. (2010). E-recruitement: recherche de mots-clés pertinents dans le titre des annonces d'emploi. 10ième Journées internationales d'Analyse statistique des Données Textuelles, pp. 4-7.

Sieg, A., Mobasher, B., & Burke, R. (2007, Novembre). Learning Ontology-Based User Profiles: A Semantic Approach to Personalized Web Search. IEEE Intelligent Informatics Bulletin, 8(1), 7-18.

Sklab, Y., Nacer, H., & Tari, A. K. (2013). Modeling User Context for Pervasive Computing. *Proceedings of the International Workshop on Requirements Engineering,* Constantine, Algeria (pp. 40-46). .

Tamine, L., Zemirli, N., & Bahsoun, W. (2007). Approche statistique pour la définition du profil d'un utilisateur de système de recherche d'information. In Information - Interaction – Intelligence, Cépaduès Editions, 7, pp. 5-25.

Torres, M. J. M., & Rodrigez, J. R. (2010). Reg: Un algorithme glouton appliqué pour le résumé automatique. *Proceeding of 10th International Conference on Statistical Analysis of Textual Data*, Montreal, Canada.

Weib, D., Scheuerer, J., Wenleder, M., Erk, A., Gûlbahar, M., & Popien, C. L. (2008). A user profile-based personalization system for digital multimedia content. *Proceedings of the DIMEA'08 proceeding of the 3rd international conference on Digital Interactive Media in Entertainment and Arts* (pp. 281-288). ACM.

ENDNOTES

[1] http://nlp.stanford.edu/software/stanford-dependencies.shtml

Taous Iggui is a PhD student in Computer Science. She is in the second year of her doctoral program at LIMED Laboratory, University of A. Mira, Bejaia, Algeria. She received her Master's degree in Networks and Distributed Systems from the same institution (2013). Her main interests are related to Data Mining, Text Processing, Knowledge Management, Semantic Web and User Profiling.

Hassina Nacer received an engineering degree in Computer Science from ESI Algiers, a Magister degree in Computer Science from Algiers University, and a PhD thesis in distributed networks from Bejaia University. She worked as an engineer at the Algerian customs for 6 years. She has served as a teacher-researcher in the Department of Computer Science at Bejaia University since 2001. In 2015, she joined the Department of Computer Science at University of Science and Technology "USTHB", Algiers, as an Assistant Professor. She has written several papers on Semantic Web services Technology and Distributed Systems. Her research interests include Knowledge Base Systems, Distributed and Cooperative Computing, Semantic Web, Web Services and Web Services Security.

Youcef Sklab is a Lecturer within the Computer Science Department in Bejaia University. He is a member of the LIMED Laboratory of the same university. He received both his Engineer's degree and Magister in Computer Science (2013) from the same institution. He is particularly interested in the modeling and managing the user context to support monitoring of user behaviors and activities.

Taklit Ait Radi received a Master's degree in Computer Science from Bejaia's University (2013).