

Text Template Mining using Named Entity Recognition

Dania Sagheer
PhD Student
Dept. Artificial Intelligence
Faculty of Informatics Engineering, Aleppo
University

Fadel Sukkar
Professor
Dept. Artificial Intelligence
Faculty of Informatics Engineering, Aleppo
University

ABSTRACT

In this paper, the named entity recognition system is built using morphological, lexical and semantic analysis. Rule based system is designed for template mining from the Arabic text. Arabic texts are selected from oil production domain. They are taken from Arabic BBC, RT and CNN websites. The System is tested on these texts and the results give high performance, less error made and good accuracy in finding the templates from texts according to named entities extracted.

Keywords

Named Entity Recognition, Template Mining, Morphological Analysis, Lexicon, Semantic Analysis.

1. INTRODUCTION

Digital documents are spread in the net as political, economic and financial news, cooking recipes, and posts on social webs. All these texts are full of information. For each domain, the texts include different types of information. For example, in social websites, texts contain information about the artists and cooks; whereas texts in economic news contain information about organizations and products. Most people do not have the time and the ability to read full text, so the need to build automatic systems is necessary to give end users the requested information without reading the full text. These systems extract useful information from texts to help end users.

Named entity recognition is the first step in information extraction systems because a named entity recognition system can detect and extract words which are classified to the specific named entity [1]. Named entity categories are determined according to specific domains.

Template mining is important for learning information about the text [2]. Many tasks of the Natural Language Processing as summarization and question answering system depend on extracted templates [3]. Template based system deals with a specific domain and is used for web search engines [4]. Designed templates are related to the task which needs these templates [5]. Many of template based systems are built, some of them depends on previously designed templates such as [6], whereas other don't have the templates previously and depend on methods as rules to mine the text templates [7]. In [8] the templates are mined from a plain text corpus for finding paraphrases in text using part of speech tagger to extract entities. These entities are used for the slots of templates. Knowledge is extracted from uninstructed texts [9] by identifying named entities and finding relationships between these entities to form templates.

Named entity recognition is important in semantic processing of text, and makes a challenge task. [10]

In this system the selected domain is oil production, so the named entities are related to finance and production. The

templates are not designed previously, the goal of the proposal system is to min and extract the templates from the text.

The named entities are determined by named entity recognition system. Named entity recognition system depends on Al-Khalil Morphological Analyzer [11], and on the Stanford Model [12]. Recognized named entities are entered to the Rule Based System to min the appropriate templates. Rule Based System uses Al Ramus lexicon [13].

The paper shows in the second section Preparation of text, the third section presents the named entity recognition system divided into morphological analyses and semantic analyses. The fourth section explains the template mining consisted of Rule Based System. The results are presented in the fifth section, and the conclusion is displayed in the sixth section.

2. PROCESSING OF TEXT

The texts are gathered from the Arabic BBC, Arabic CNN, RT Arabic newspapers. The texts are split into paragraphs according to conjunction marks as points.

3. NAMED ENTITY RECOGNITION SYSTEM

3.1 Morphological analysis

The text is analyzed morphologically using Al Khalil lexicon. Roots and stems are taken from Morphological Analyzer. The names of entities are related to oil production domain. They comprise thirteen entities such as "organization", "منظمة", "time", "زمن", "unit of oil", "وحدة قياس النفط", "currency", "عملة", "level", "مستوى".

Morphological analyzer has a dataset. The dataset contains verbs, nouns, roots, proper nouns, stop words, prefixes and suffixes. Morphological analyzer removes the prefixes and suffixes to get the stem depending on nouns in the dataset. Morphological analyzer also removes extra letters and gets the root by checking the output after removing the roots in the dataset.

Stanford model uses corpus containing closed statues of words as prepositions, non-derivative words, and open statues of words as nouns and verbs which can be derived and is trained to give the Part of Speech of the words.

This System uses Al-Khalil morphological analyzer and Stanford Model to analyze the text and get the morphological statues of these words.

The users don't know the named entities and the system has to access the text paragraph by other to identify each word as a named entity or unnamed-entity.

- The named entity “verb”, “فعل” is determined using AL Khalil analyzer and Stanford model. Al Khalil depends on patterns and verbs saved in the dataset. Stanford model gives parts of speech (POS) of Arabic words.
- If the analysis of Al-Khalil morphological analyzer on this word is “verb”, “فعل” and the POS tag that this word is a verb, the word will be a named entity called “Verb”, “فعل”.

3.2 Semantic Analyzer

The semantic analysis depends on the word definition in the Al Ramus lexicon.

Al Ramus lexicon contains verbs, nouns, morphological statuses and definitions for each word. The definition expresses the semantic meaning of the word.

- If a word contains in its definition in Al Ramus “oil producing organization”, “منظمة منتجة للنفط” it will be a named entity called “organization”, “منظمة”.
- If a word contains in its definition in Al Ramus “political person”, “شخصية سياسية” it will be a named entity called “political person”, “شخصية سياسية”.
- If a word contains in its definition in Al Ramus “level”, “مستوى” it will be a named entity called “level1”, “مستوى 1”. In this time the system search about word which has root equals “price” “سعر” or has stem equals “production” “إنتاج” this word will be a named entity called “level 2”, “مستوى 2”.
- If a word contains in its definition in Al Ramus “underground wealth”, “ثروة باطنية” it will be a named entity called “underground wealth”, “ثروة باطنية”.
- If a word contains in its definition in Al Ramus “coin unit”, “وحدة نقدية” it will be a named entity called “currency”, “عملة” then, get the previous word to check whether if it’s a number to extract named entity called “number” “عدد”.
- If a word contains in its definition in Al Ramus “country”, “دولة” it will be a named entity called “country”, “دولة”.
- If a word contains in its definition in Al Ramus “normal number”, “عدد طبيعي” it will be a named entity called “numerical statue”, “منزلة رقمية”. Then check the previous word whether if it’s a number to extract named entity “number”, “عدد”.
- If a word contains in its definition in Al Ramus “oil measure”, “قياس النفط”, it will be a named entity called “unit”, “وحدة”, also the previous word is checked whether if it’s a number to detect named entity “number”, “عدد”.
- If the morphology state of a word is rigid and this word contains in its definition in Al Ramus “month”, “شهر” or “day”, “يوم” it will be a named entity called “time”, “زمن”.
- If a word contains in its original in Al Ramus “related to”, “منسوب إلى” and stem of this word, the stem has to be country name, this word will be a named entity called “related to country”, “منسوب إلى دولة”.

4. TEMPLATE MINING

4.1 Rule Based System

The recognized named entities are entered to the rule based system to extract text templates.

The templates will be extracted from the text by general rules. The text is split into paragraphs. The templates are mined from each paragraph.

The rules are:

1. If paragraph named entities contain “Political Person”, “شخصية سياسية”, the system doesn’t care about the entity “Related to Country”, “منسوب إلى دولة”.
2. If paragraph entities contain “verb”, “فعل”, “political person”, “شخصية سياسية”, “underground wealth”, “ثروة باطنية”, “organization”, “منظمة”, “level1”, “مستوى 1” and “level2”, “مستوى 2”, the first template is created.
3. If paragraph entities only contain “Underground wealth” “ثروة باطنية” and “organization” “منظمة” not found, no template is mined from this paragraph because it isn’t decision about an event as the level of oil production and the underground wealth not belong to any organization.
4. If paragraph entities only contain “organization” “منظمة” and don’t contain entity “political person” “شخصية سياسية” or entity “underground wealth” “ثروة باطنية”, no template in this paragraph is extracted, because the organization doesn’t present any information about the persons or about the levels of underground wealth production.
5. If paragraph entities contain “verb”, “فعل”, “political person”, “شخصية سياسية”, level1 “مستوى 1”, “level2” “مستوى 2” and “time” “زمن”, the second template is mined, so the political person decides information related to the level and done in this time.
6. If paragraph contains the entities “verb” “فعل”, “underground wealth” “ثروة باطنية”, “level1” “مستوى 1”, “level2” “مستوى 2” and “time” “زمن”, the third template is extracted, this template expresses about the level of underground wealth without attention to the political person or organization.
7. If the named entities of the paragraph contain “Organization” “منظمة”, the entity “country” “دولة” isn’t cared.
8. If paragraph contains entities “verb” “فعل”, “organization” “منظمة”, “level1” “مستوى 1”, “level2” “مستوى 2”, “numerical statue” “منزلة رقمية”, “number” “عدد”, “unit” “وحدة” and “time” “زمن”, the fourth template is mined. This template determines the level by numerical value and measurement unit in a detected time.
9. If paragraph contains entities “verb” “فعل”, “organization” “منظمة”, “level1” “مستوى 1”, “level2” “مستوى 2”, “numerical statue” “منزلة رقمية”, “number” “عدد”, “currency” “عملة”, and “time” “زمن”, the fifth template is mined, this template similar to the fourth template but here the level is measured by the currency not with the unit of production.

Figure 1 shows the block diagram of the system for template mining of text.

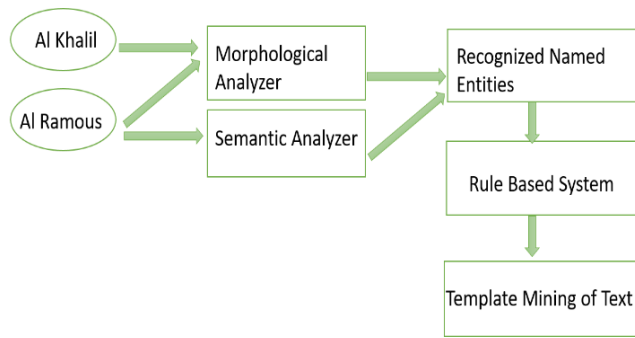


Fig 1: Block Diagram for template mining of text

5. RESULTS

5.1 First Text (BBC 7/12/2018)

Text in Arabic

“ارتفعت أسعار النفط عقب اتفاق منظمة الدول المصدرة للنفط أوبك على خفض الإنتاج بأكثر من 1 مليون برميل يوميا في محاولة لتعزيز الأسعار.....

وجاءت الخطة التي تتضمن خفض الإنتاج بواقع 1 مليون برميل يوميا استجابة للتراجع الحاد في أسعار النفط العالمية بحوالي 30 في المئة منذ أكتوبر تشرين الأول الماضي.....

وتتضمن الخطة خفض الدول الأعضاء في منظمة أوبك إنتاجها في فترة الاتفاق بواقع 800 ألف برميل يوميا بينما تخفض الدول الحليفة لأوبك إنتاجها بواقع 400 ألف برميل يوميا.

وقالت كارولين باين خبيرة الأسواق: "نتوقع أن يستمر تباطؤ النمو الاقتصادي في التأثير في الطلب على النفط العام المقبل وهو ما يتزامن مع نمو سريع في إنتاج النفط الأمريكي وهو ما سيؤثر على الأسعار".

Text in English:

“Oil prices soared after the Organization of Petroleum Exporting Countries (OPEC) agreed to cut production by more than 1 million barrels per day in an effort to boost prices

The plan, which includes a cut in production by 1 million barrels per day in response to the sharp decline in world oil prices by about 30 percent since October last

The plan includes cuts by OPEC member states in the period of the agreement by 800 thousand barrels per day, while the OPEC countries reduce production by 400 thousand barrels per day.

"We expect the slowdown in economic growth to continue to affect oil demand next year, which coincides with rapid growth in US oil production, which will affect prices," said Caroline Pine, a market expert."

The text is split to get its paragraphs according to the point mark. As in the following table 1.

Table 1. Paragraphs of First Text

#	Paragraph
1	"ارتفعت أسعار النفط عقب اتفاق منظمة الدول المصدرة للنفط أوبك على خفض الإنتاج بأكثر من 1 مليون برميل يوميا في محاولة لتعزيز الأسعار" "Oil prices soared after the Organization of Petroleum Exporting Countries (OPEC) agreed to cut production by more than 1 million barrels per day in an effort to boost prices"
2	"وجاءت الخطة التي تتضمن خفض الإنتاج بواقع 1 مليون برميل يوميا استجابة للتراجع الحاد في أسعار النفط العالمية بحوالي 30 في المئة منذ أكتوبر تشرين الأول الماضي" "The plan, which includes a cut in production by 1 million barrels per day in response to the sharp decline in world oil prices by about 30 percent since October last"
3	"وتتضمن الخطة خفض الدول الأعضاء في منظمة أوبك إنتاجها في فترة الاتفاق بواقع 800 ألف برميل يوميا بينما تخفض الدول الحليفة لأوبك إنتاجها بواقع 400 ألف برميل يوميا." "The plan includes cuts by OPEC member states in the period of the agreement by 800 thousand barrels per day, while the OPEC countries reduce production by 400 thousand barrels per day."
.....
11	"وقالت كارولين باين خبيرة الأسواق: "نتوقع أن يستمر تباطؤ النمو الاقتصادي في التأثير في الطلب على النفط العام المقبل وهو ما يتزامن مع نمو سريع في إنتاج النفط الأمريكي وهو ما سيؤثر على الأسعار." "We expect the slowdown in economic growth to continue to affect oil demand next year, which coincides with rapid growth in US oil production, which will affect prices," said Caroline Pine, a market expert. "

For each paragraph, the system take word by other and analyzes it to recognize the correct entity for this word.

For first paragraph, the first word in Arabic is “ارتفعت” “soared”, the morphological analyzer of it returns that entity is “verb”, second word is “أسعار” “prices” morphological analyzer returns that it is a noun, but the semantic analyzer doesn’t match with any entity. Third word is “oil” “النفط”, the morphological analysis also return noun, and in the semantic analyzer the definition of this word contains “underground wealth” so the word is recognized as “underground wealth”. After getting the entity “level1” “مستوى1”, it is searched about next word, in paragraph, which match the root “price” “سعر” or stem “production” “إنتاج”. When getting this word, it will be entity “level2” “مستوى2”. such that all words in the paragraph are recognized as appropriate named entities.

Recognized Named entities of first text are showed in the following table 2.

Table 2. Recognized Named Entities for each Paragraph of First Text

#	Named Entity	Word from text
1	Organization, "منظمة"	Opec "أوبك"
	underground wealth, "ثروة باطنية"	Oil "النفط"
	numerical statue, "منزلة رقمية"	Million "مليون"
	number, "عدد"	1
	Time, "زمن"	After "عقب"
	Level1, "مستوى 1"	Reduction "خفض"
	Level2, "مستوى 2"	Production "الإنتاج"
	Verb, "فعل"	Soared "ارتفعت"
	Time, "زمن"	Daily "يوميًا"
	Unit, "واحدة"	Barrel "برميل"
4	Country, "دولة"	"United State", "الولايات المتحدة"
	Level2, "مستوى 2"	Production "إنتاج"
	Level1, "مستوى 1"	increase "زيادة"
	underground wealth, "ثروة باطنية"	Oil "النفط"
	Level1, "مستوى 1"	Fall back "تراجع"
	Verb, "فعل"	It collapsed "انهارت"
----	-----	
11	Verb, "فعل"	Continue "يستمر"
	Verb, "فعل"	Synchronize "يتزامن"
	underground wealth, "ثروة باطنية"	Oil "النفط"
	Verb, "فعل"	effect "يؤثر"

Rule Based System is applied to mine Templates from first text.

For each paragraph, recognized entities in this paragraph are entered to the rules to get the appropriate templates. in first paragraph, recognized entities correspond the eight rule so the template four is extracted, the four template focus on the quantity of the production, so when it is found paragraph entities express about quantity of organization production as "organization", "numerical statue", and "unit of oil", the four template is mined. in second paragraph, recognized entities correspond the six rule which check the entity called "underground wealth" "ثروة" to detect a level of underground wealth production. So this template care about the entities called "underground wealth", "level1", and "level2".

Extracted templates from first text are showed in the following table 3.

Table 3. Extracted Templates from First Text

Parag Num	Tem Num	Template as entities
		Templates as words from text
1	4	"Verb" "فعل" + "organization" "منظمة" + "level1" "مستوى 1" + "level2" "مستوى 2" + "number" "عدد" + "numerical statue" "منزلة رقمية" + "unit" "واحدة قياس النفط" + "time" "زمن"
		"Soared" "ارتفعت" + "Opec" "أوبك" + "reduction" "خفض" + "production" "الإنتاج" + "million" "مليون" + "1" + "barrel" "برميل" + "daily" "يوميًا"
6	4	"Verb" "فعل" + "organization" "منظمة" + "level1" "مستوى 1" + "level2" "مستوى 2" + "number" "عدد" + "numerical statue" "منزلة رقمية" + "unit" "واحدة قياس النفط" + "time" "زمن"
		"reduce" "تخفيض" + "Opec" "أوبك" + "reduction" "خفض" + "production" "الإنتاج" + "400" + "thousands" "آلاف" + "barrel" "برميل" + "daily" "يوميًا"
2	3	"Verb" "فعل" + "underground wealth" "ثروة" + "level1" "مستوى 1" + "level2" "مستوى 2" + "time" "زمن" + not("number" "عدد" + "numerical statue" "منزلة رقمية" + "currency" "عملة")
		"reach" "وصلت" + "oil" "النفط" + "reduction" "خفض" + "production" "إنتاج" + "daily" "يوميًا"

In first text, 77 words are identified as named entities, however 12 of them failed. So the error made for first text is 0.15 %.

5.2 Second Text (BBC 22/1/2017)

Text in Arabic:

"أكد وزير الطاقة السعودي خالد الفالح أن منتجي نفط من داخل منظمة الدول المصدرة أوبك وخارجها ملتزمون بتعهداتهم بخفض الإنتاج.

وأعرب الفالح عن اعتقاده بأن المخزون العالمي من النفط قد يعود إلى متوسط خمس سنوات بحلول منتصف العام الجاري مع الأمتثال الكامل بالخفض.

وقال الوزير السعودي: " لا يوجد ما يدعونا للقول على نحو مفاجئ مع دخول يناير أننا بحاجة إلى خفض أكبر أو فترة أطول هل هذا سيتغير في الربع الثاني هذا ممكن لكنه غير مناسب اليوم".

وأوضحت موسكو أنها ستخفض الإنتاج بواقع 300 ألف برميل بحلول أبريل نيسان أو مايو أيار حسب ما ذكرت شبكة بلوم بيرغ الأمريكية.

Text in English:

"Saudi Oil Minister Khalid al-Falih said oil producers from within and outside the Organization of Petroleum Exporting Countries (OPEC) are committed to cutting production.

Al-Faleh believed that global oil reserves could return to an average of five years by the middle of this year with full compliance with the reduction.

For second text, 109 words are recognized as named entities, however 15 of them are not identified correctly. So the error made for second text is 0.13 %.

5.2 Third Text (RT Arabic 13/2/2019)

Text in Arabic:

"أفاد وزير الطاقة الروسي ألكسندر نوفاك الثلاثاء بأنه من المتوقع انخفاض إنتاج روسيا من الخام خلال الشهر الجاري بواقع 90 – 100 ألف برميل يوميا عن مستواه في أكتوبر الماضي .

وتأتي هذه الخطوة في إطار اتفاق عالمي باسم أوبك مبرم بين منتجين من داخل وخارج أوبك ينص على تقليص الإنتاج بواقع 1 مليون برميل يوميا ابتداء من 2019 بهدف تحقيق توازن بين العرض والطلب في السوق النفطية.

من جهتها أعلنت منظمة الدول المصدرة للنفط أوبك الثلاثاء أنها قلصت إنتاجها في يناير الماضي إلى 30 مليون برميل يوميا وذلك تماشيا مع الاتفاق.

ووفقا لتقرير المنظمة الشهري فقد كانت حصة الأسد من خفض الإنتاج للسعودية التي قلصته بمقدار 350 ألف برميل في اليوم تلتها في ذلك الإمارات ثم الكويت"

Text in English:

"Russia's crude production is expected to fall by 90-100 thousand barrels per day from its level in October, Russian Energy Minister Alexander Novak said Tuesday.

The move comes within the framework of a global agreement called "OPEC +," concluded between producers from within and outside the "OPEC", which provides for the reduction of production by 1.2 million barrels per day starting in 2019 in order to achieve a balance between supply and demand in the oil market.

The Organization of Petroleum Exporting Countries (OPEC) said on Tuesday that it cut production in January to 30.81 million barrels per day, in line with the agreement.

According to the organization's monthly report, the lion's share of production cuts to Saudi Arabia, which was cut by 350 thousand barrels per day, followed by the UAE and Kuwait"

After splitting the text into paragraphs, the named entities are got in the following table 7

Table 7: Named Entities in Third Text

#	Named Entity	Word from text
1	Level1 "مستوى 1"	Reduction "انخفاض"
	Time "زمن"	daily "يوميًا"
	Verb "فعل"	attend "شهد"
	Number "عدد"	100
	Political person "شخصية سياسية"	Minister "وزير"
	Numerical statue "منزلة رقمية"	Thousand "ألف"
	Level2 "مستوى 2"	Production "إنتاج"
	Country "دولة"	Russia "روسيا"
	Verb "فعل"	benefit "أفاد"
2	Unit "واحدة"	Barrel "برميل"
	Verb "فعل"	include "ينص"

Number "عدد"	1
Organization "منظمة"	Opec "أوبك"
Numerical statue "منزلة رقمية"	Million "مليون"
underground wealth "ثروة باطنية"	Oil "النفطية"
Time "زمن"	daily "يوميًا"
Unit "واحدة"	Barrel "برميل"
Level2 "مستوى 2"	Production "الإنتاج"
Level1 "مستوى 1"	Reduction "تقليص"
Country "دولة"	Inside "داخل"
-----	-----

As previous, the templates of third text are extracted as in the table 8.

Table 8. Some of Extracted Templates

Parag Num	Tem Num	Template as entities
		Templates as words from text
1	4	"Verb" "فعل" + "organization" or "country" + "level1" + "منظمة أو دولة" + "level2" + "مستوى 2" + "number" + "عدد" + "numerical statue" + "منزلة رقمية" + "unit" + "واحدة قياس النفط" + "time" + "زمن"
		benefit "أفاد" + "Rashia" "روسيا" + "reducing" + "انخفاض" + "production" + "thousands" + "100" + "إنتاج" + "يوميًا" + "daily" + "برميل" + "barrel" + "ألف"
2	4	"say" "ينص" + "Opec" "أوبك" + "reduction" + "خفض" + "production" + "million" "مليون" + "1" + "إنتاج" + "يوميًا" + "daily" + "برميل" + "barrel"

In third text, the system can detect 39 words of named entities, failure 7 words, so the error made is 0.17%.

6. CONCLUSION

Template mining of Arabic text is done using named entity recognition system which gives high results and less errors. The named entities are recognized by morphological and semantic analyzers. Morphological and semantic analyzer depend on the Al-Khalil, Al-Ramus, and Stanford model to give the named entities.

The recognized named entities are entered to the Rule Based System for extract templates. Text template mining is satisfied successfully without previously designed templates and the results give high performance.

7. REFERENCES

- [1] Foley John., Sarwar M Sheikh., and Allan James., “Named Entity Recognition with Extremely Limited Data” arXiv:1806.04411v2 [cs.IR] 13 Jun 2018.
- [2] Vargas-Vera Maria, Domingue John, Kalfoglou Yannis, Motta Enrico and Shum Buckingham Simon, “Template-Driven Information Extraction for Populating Ontologies”, Proceedings of the 2nd International Conference on Ontology Learning - Volume 38, Pages 35-40, Seattle, USA, 2001.
- [3] Desai Prashant G., Sarojadevi H, Niranjana N. Chiplunkar, “A Template Based Algorithm for Automatic Summarization and Dialogue Management for Text Documents”, IJRET: International Journal of Research in Engineering and Technology, Volume: 04 Issue: 11 | Nov-2015
- [4] Gobinda G. Chowdhury, “Template Mining for Information Extraction from Digital Documents”, Division of Information Studies, School of Applied Science, Nanyang Technological, Nanyang Avenue, Singapore 639798 LIBRARY TRENDS, Vol. 48, No. I, Summer 1999, pp. 182-208.
- [5] Onyshkevych Boyan, “Template Design for Information Extraction”, Proceedings of the 5th conference on Message understanding, 1993.
- [6] Nathanael Chambers and Dan Jurafsky. Template-Based Information Extraction without the Templates, Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, 2011.
- [7] Sagheer Dania, Sukkar Fadel, “A Template-Based Information Extraction System for Text Understanding”, International Journal of Computer Applications, Vol 182, NO 28, 2018.
- [8] Biran, Or, Terra Blevins and Kathleen McKeown. “Mining Paraphrasal Typed Templates from a Plain Text Corpus.” *ACL* (2016).
- [9] Raymond J. Mooney and Razvan Bunescu, “Mining Knowledge from Text Using Information Extraction”, ACM SIGKDD Explorations Newsletter - Natural language processing and text mining, Volume 7 Issue 1, June 2005, Pages 3-10
- [10] Girish Keshav Palshikar, “Techniques for Named Entity Recognition: A Survey”, chapter 11, IGI Global, 2012.
- [11] BOUDCHICHE Mohamed, MAZROUI Azzeddine, OULD Bebah Mohamed Ould Abdallahi, LAKHOUAJA Abdelhak, BOUDLAL Abderrahim. AlKhalil Morpho Sys 2: A robust Arabic morpho-syntactic analyzer, Journal of King Saud University – Computer and Information Sciences, 2017, 141–146p.
- [12] The Stanford Natural Language Processing Group: <https://nlp.stanford.edu/software/tagger.shtml>, Access Date 1/1/2018.
- [13] Arramooz AlWaseet: Arabic dictionary for morphology: <http://arramooz.sourceforge.net>. Access Date 1/1/2018.