Journal of King Saud University – Computer and Information Sciences 29 (2017) 212–228

Contents lists available at ScienceDirect

Journal of King Saud University – Computer and Information Sciences

journal homepage: www.sciencedirect.com

Deriving ontological semantic relations between Arabic compound nouns concepts

Imen Bouaziz Mezghanni *, Faiez Gargouri

MIR@CL Laboratory, Sfax University, Tunisia

ARTICLE INFO

Article history:
Received 6 November 2016
Revised 27 February 2017
Accepted 1 March 2017
Available online 10 March 2017

Keywords:
Semantic relations derivation
Arabic compound nouns
Compound structure
Gamification
FCA
RCA

ABSTRACT

Legal ontologies have proved their increasingly substantial role in representing, processing and retrieving legal information. By using the knowledge modeled by such ontologies in form of concepts and relations, it is possible to reason over the semantic content of legal documents. Supporting (semi-) automatically the development of ontologies from text is commonly referred to as ontology learning from text. The learning process includes learning of the concepts that will form the ontology and learning of the semantic relations among them.

In this paper, we present a new approach for expliciting the semantic relations between Arabic compound nouns concepts. The originality of this work is twofold. Firstly, the technique of inferring relations is based on exploiting the internal structure of the compounds using a defined set of domain-and language-independent rules according to their different structures, on the one hand, and on studying prepositions semantics specifying the inferred relations applying a gamification mechanism that collects human votes, on the other hand. Secondly, relying on the compounds set described by both binary (structural positions in which there are written) and relational attributes (the deduced relations), we used a "Relational Concept Analysis" (RCA) technique, as an adaptation of "Formal Concept Analysis" (FCA), for the construction of interconnected lattices that we transformed into ontological concepts and relations which can be either taxonomic or transversal.

Experiments carried out on Arabic legal dataset showed that the proposed approach reached encouraging performance through achieving high precision and recall scores. This performance affects positively the retrieval results of legal documents based on a powerful ontology, which presents our main objective.

1. Introduction

With the development of information technology and easier internet access, electronic dissemination of huge amounts of published documents has made the legal information retrieval more and more complex for the user. Nowadays, search engines present the main tools for accessing data available on the Web. However, most search engines do their text query and retrieval using keywords, which often results in hits completely irrelevant to user query leading to low precision and recall parameters. The weakness of search engines can be overcome through using Semantic Web technologies considered to be the next generation of the actual Web. The Semantic Web is a Web of ontologies that allow the analysis of the domain knowledge by modeling the relevant concepts of this domain. The ontologies enable semantic interoperability involving the comprehension of information to be precisely described and well understood by machine. Therefore, the search is no longer based on keywords matching, but rather on concepts matching. In this case, the search results become more relevant, which increases precision and recall rates.

However, the manual building of ontologies is a time consuming and labor intensive task. Ontology learning (Maedche and Staab, 2004), which aims at providing automatic and semi-automatic approaches for ontology generation, can overcome the bottleneck of knowledge acquisition. The learning process includes learning of the concepts that form the ontology and learning of the semantic relations among them. This paper introduces a novel approach for expliciting the semantic relations between Arabic compound nouns concepts.

To further explain the proposed approach, it is necessary to define the following terms:

* Corresponding author.
E-mail address: imen_bouaziz_miracl@yahoo.com (I.B. Mezghanni).

Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.jksuci.2017.03.001
1319-1578/© 2017 The Authors. Production and hosting by Elsevier B.V. on behalf of King Saud University.

- "A term is a lexical unit consisting of one or more than one word which represents a concept inside a domain" (de Bessé et al., 1997).
- "A concept is an abstract unit which consists of the characteristics of a number of concrete or abstract objects which are selected according to specific scientific or conventional criteria appropriate for a domain" (de Bessé et al., 1997).
- "A multi-word term or compound term is a combination of a set of words used to convey a single unit of meaning. Its semantics depends on the knowledge area of the concept it describes and cannot be inferred directly from the semantic composition of its components separately" (Sag et al., 2002).

The association "term = concept" is erroneous. Indeed, a term can represent many concepts. For example, the term "draft" can refer either to a current of air into an enclosed space or to the first version of a document, plan or drawing. However, a concept may be denoted by many terms. Therefore, terms are considered as units of language, while concepts are elements of the conceptual model.

Whatever the text and the language in which the compounds are written, they are often viewed as relevant since they play an important role in the encapsulation and expression of nominal concepts. Compounds are also frequent in a wide variety of texts types, which makes their extraction a crucial task.

In a conceptual model, considering compound concepts without taking into account a predefined relation linking them is not very significant as it may lead to their discard. Determining the semantic relations between concepts is fundamental in capturing the ideas in texts. Besides, relations, such as part-whole, cause-effect, etc., encode crucial information about how different entities should be perceived in relation to each other. Thus, much attention has been paid to this research field and several works have recently been carried out on different languages, such as English (Ta and Thi, 2016; Joseph et al., 2016) and Chinese (Miao et al., 2012), etc.

For instance, extracting semantics from compound nouns was tackled by Vela and Declerck (2009) in a process of ontology building. Relying on pattern-based approaches, compounds were first detected and analyzed to suggest candidate ontology classes and relations. Then, paraphrases of the compounds in the text were detected through a set of patterns and analyzed in order to filter and validate the list of candidate classes and relations obtained in the first step. In their approach, only noun-noun compounds were taken into account.

With the intention of automatic Thai ontology construction, Kawtrakul et al. (2004) processed parses sentences and generated compound nouns as candidate terms based on phrase chunking. Using statistical-based technique, the compounds were analyzed in order to separate head and modifier from each other. The semantic relations of a compound were extracted by learning the common ancestral concept of its head and modifier using heuristic rules as well as expert's judgments.

Sruti Rallapalli (2012) explored the scope of identifying the semantic relation. Thereby, he interpreted compound nouns using an indexed semantic ontology combined with noun similarity measurement techniques. The problem, here, is that the semantic similarity is limited to the ontology itself as the primary information source. Therefore, there is a need for creating standard corpora in any domain of application.

Extracting semantic relations from compound nouns can be also based on a frame-semantic approach (Lakhfif and Laskri, 2016). The basic idea of the latter is that meanings, such as purpose, constitution and agency, their realization,etc., can be viewed as a generalized and lexicalized aspect of qualia structure as defined by Pustejovsky (1991). In this context, the challenge consists in the ability to organize relational possibilities hierarchically according to the compounds underlying semantic meanings and the ability to recognize an implication structure among different but related relational possibilities.

However, despite the importance of the Arabic language, few studies investigated the process of extracting the semantic relations in Arabic texts due to the complexity of this task. This complexity arises from the distinctive features characterizing the Arabic language, namely the agglutination and diactritization causing major morphological and syntactic ambiguities.

In this paper, the derivation of semantic relations between Arabic compound nouns is dealt with through developing a hybrid approach to combine the advantages of clustering and rule-based approaches. From the compounds, two kinds of implicit relations (is-a relation; objectProperty relation) are extracted based on a set of pattern rules defined according to the different structures of the compounds. To specify the resulting objectProperty relation, we resorted to prepositions in order to describe the hidden relations present in the compounds through a gamification mechanism. Gamification refers to "the use of design elements characteristic for games in non-game contexts" (Deterding et al., 2011). A validation step was followed by experts in order to verify the accuracy of the chosen relations and the reliability of the proposed rules.

This work introduces a part of ontology construction whose goal is to support retrieval of legal documents and in which we focus on the structural positions where the concepts appear in the document. This position is determined by referring to such structural element of the document. In a legal code, we considered a structural position, the article number to which a concept belongs (Mezghanni and Gargouri, 2015). Thus, a concept is described by the "Articles" where it is written. For instance, the concept حاكم التحقيق (investigating judge) is described by article 10 and article 11. The ontological concepts together with their associated positions are defined by means of an incidence matrix to FCA (Ganter and Wille, 1997; Ganter et al., 2005) which is a mathematical approach for data analysis providing a rigorous framework for the derivation of a conceptual hierarchy called "concept lattice".

In order to handle the generated relations between concepts (the concept حاكم التحقيق (investigating judge) is-a حاكم (judge)), we relied on RCA (Huchard et al., 2007, 2003) as an extension of FCA which includes further relational structures. Indeed, RCA considers the relations between objects in addition to the characteristics of the objects (sets of object-attribute data provided with relations). In other words, objects are described by attributes and their relations with other objects. RCA consists in iteratively applying an FCA algorithm using relational data. The discovered concepts at a given step are propagated along the relations, leading to the discovery of new concepts at the next iteration.

The remainder of the paper is organized as follows: Section 2 discusses recent works in the domain of semantic relation extraction from Arabic texts. Then, we recall the basic notions of ontologies, FCA and RCA in Section 3. The adopted approach is described in Section 4. Section 5 shows the experiments and the obtained results evaluated in Section 6. A conclusion with future research directions are presented at the end of the paper.

2. Related work

In the literature, several researches were conducted to investigate the process of Arabic ontologies learning in different applications. These ontologies belong to various domains and were constructed differently. The survey proposed in Mezghanni and Gargouri (2015) summarizes recent works presented for ontology learning from Arabic textual resources.

Besides, automatic extraction of relations either ontological or not has attracted many researchers in different languages such as English (Xiang et al., 2016; Devisree et al., 2016). However, little attention has been paid to Arabic. In general, these works can essentially be classified according to the used extraction approach into four main categories: rule-based approaches, clustering-based approaches, machine learning-based approaches and hybrid approaches.

2.1. Rule-based approaches

These approaches are based on patterns embedding all the potentially-related linguistic sequences usually implemented in the form of regular expressions or finite-state transducers.

Sadek and Meziane (2016) extracted causal relations that are explicitly expressed in Arabic texts based on a pattern recognizer model. This model incorporated a set of around 700 linguistic patterns allowing the distinction of the sentences parts which represent the cause and the effect. The patterns were generated based on different sets of syntactic features by analyzing a large untagged Arabic corpus.

Although such approaches are very interesting for a restricted domain and have a good analysis quality, they cannot perform well, especially that the process of manually hand-crafting patterns is too expensive in terms of time and effort. Thus, by applying these approaches, it is hard to process large quantities of data.

2.2. Clustering-based approaches

In the clustering-based approaches, each cluster of entity pairs is likely to contain semantic variations of the same relation. The relation between two entities is defined by their context which includes a set of features varying from entity semantic information to lexical and syntactic features in all the co-occurrence of entities. These contextual features can represent relation between entity pairs. Therefore, the labels used to describe the relations among entities are extracted from the clustering process result. For instance, FCA is a popular conceptual clustering approach employed for hierarchical relations extraction. Indeed, this technique has not been yet applied on Arabic texts.

To entirely automate the relation extraction task, some research studies adopted machine learning approaches including unsupervised, semi-supervised and supervised techniques.

2.3. Machine learning-based approaches

In the unsupervised methods, a common approach builds clusters of patterns that express the same relation and generalizes them. However, due to the semantic meaning representation of relational patterns and scalability to large data, it is challenging to obtain a reliable set of patterns (Takase et al., 2015). Although these approaches can process very large amounts of data, the mapping of the resulting relations to ontologies is quite hard. To the best of our knowledge, no work has dealt with Arabic relation extraction using unsupervised techniques.

To overcome the problems encountered by unsupervised methods, studies have recently relied on semi-supervised techniques or bootstrapping methods that require only a small set of seeds instead of a training set. These seeds can be depicted as a sample of linguistic patterns or some target relation instances to acquire more basics until finding all target relations in an iterative way. The disadvantage of the bootstrapping methods depends deeply on the selected initial seeds that must accurately reflect the information presented in the corpus. Otherwise, the quality of extractions might be low.

Accordingly, even though the results of these methods are very promising, they suffer from low precision affected by the error propagation caused by the incorrect or too general used patterns. Since the semi-supervised methods require several iterations, these methods are prone to semantic drift (such as an unwanted shift of meaning). This signify that these methods require a certain amount of human effort to create seeds initially and to help keep systems “on track” to prevent them from semantic drift (Augenstein et al., 2014).

In fact, Al-Yahya et al. (2016) faced this problem while developing “Badea system” designed to the semi-automated enrichment of ontological lexicons. They used a pattern-based approach using a seed ontology, composed of a small set of antonym pairs, to extract pairs of words from a given corpus with the antonym semantic relation. Then, the discovered pairs are used to enhance the ontology. In order to avoid this problem and improve the precision score, Al-Yahya et al. (2014) extended their above-mentioned work through employing LogDice score and calculating a score for each pattern based on its co-occurrence.

Al Zamil and Al-Radaideh (2014) improved Hearst algorithm (Hearst, 1992) proposed to detect automatically hyponyms by constructing lexical patterns of knowledge. In order to overcome the main disadvantage of this algorithm consisting in its large human intervention requirement for the creation of patterns from real examples, the enhancement process was carried on by generating a system designed to analyze Arabic text using lexical semantic patterns according to a set of features for the extraction of ontological relations. But, the major problem of their approach is that the detection of frequent classification errors affects negatively the overall performance of the proposed technique.

The third approach depends on a binary classification task in which a classifier is trained using a set of either negative or positive examples of specific semantic relations. As it requires a large fully-labeled corpus, using such approaches in different domains necessitates more manual effort.

Boujelben et al. (2014b) proposed a relation extraction system named “RelANE” that discovers the semantic binary relations between Arabic named entities. In their system, for each word in the sentence, a set of morphological, contextual and semantic features of entity types was used. Nonetheless, RelANE has two main drawbacks. Firstly, many relations were not extracted due to the incorrect POS tags and the declassification of the named entities. Moreover, the evaluation was performed on a manually constructed data set corpus instead of other available dataset, such as the free ANERCorp,[1] the commercial ACE[2] and ALTEC.[3]

One year later, Falih and Omar (2015) proposed an Arabic grammatical relation extraction approach. Its main objective is to label each Arabic word with the corresponding grammatical relation (subject, object or predicate). The score achieved by this approach was better than that reached by RelANE system. However, its disadvantage was also in the evaluation phase, as it is in the case (Boujelben et al., 2014b), since a small manually-created corpus was used with only 80 sentences, which might lead to an unfair evaluation.

2.4. Hybrid approaches

Recently, many researchers have tried to combine these approaches in the so-called hybrid approaches to obtain better results. Indeed, to enhance the performance of the proposed approach, it is better to combine the pattern-based approach and machine learning-based approach than to use each method sepa-

[1] http://users.dsic.upv.es/~ybenajiba/.

[2] http://catalog.ldc.upenn.edu/LDC2006T06.

[3] http://www.altec-center.org/Repository_65.html.

rately. For instance, benefiting from pattern-based algorithms only is very difficult since these algorithms often require bootstrapping or initial clustering, which can be done through machine learning methods. Moreover, machine learning-based approaches can be combined with pattern-based approaches to prevent bad results due to the lack of knowledge and lack of precision.

Boujelben et al. (2014a) adopted a hybrid method to extract relations between Arabic named entities. The authors built a linguistic and learning model to predict the positions of words which express a semantic relation within a clause. The approach employed linguistic modules to ameliorate the results provided by using a machine learning-based approach. The achieved performance was encouraging. The empirical results indicated that the hybrid approach outperformed both the rule-based system and the machine learning-based approaches in terms of the F-score when applied to the same standard testing data set, ANERCorp. Although it has a promising performance, the process cannot extract some of the relations between words that are not close to the named entities' positions, notably in the case of long and the complex sentences.

2.5. Motivation

Most of the previously-mentioned works capture only the explicit semantic relation between terms without considering the implicit ones despite their importance, which affects relatively the accuracy of their evaluation.

In this paper, we propose a new approach of explicit and implicit semantic relations derivation. This work fits into the context of ontology construction from Arabic texts applied to the legal field. The first step for the construction process is the extraction of relevant concepts (simple and compounds) through a hybrid approach combining a pattern-based approach and a learning algorithm. In this paper, we are not really interested in how the concepts are extracted as it is the case in Mezghanni and Gargouri (2015). Nevertheless, we give a short overview on the process.

In fact, this step is principally based on NooJ platform[4] through which we elaborated three different types of grammars. A morphological grammar was built for decomposition of Arabic agglutinated words. An inflectional/derivational grammar was developed to generate the different voweled forms of the dictionary entries as well the grammatical variants of the same word. Two syntactical grammars were also created. The first one was used to extract the logical structure of the documents, while the purpose of applying the second was to extract all related derived and agglutinated forms. This grammar is composed of 19 sub-graphs; each of which contains the appropriate processing of the specific grammatical category. The projection of these grammars on corpus resulted in annotated documents considered as input to a learning algorithm relying on various features classified into structural, content and semantic ones to keep the relevant concepts.

The second step, the focus of the present paper, is the detection of relations connecting the already extracted concepts.

Unlike the existing relations extraction approaches, ours follows the below-mentioned steps:

- the inference of general relations based on the rules defined according to the internal structure of the compounds concepts,
- the inference of specific relations by finding the most likely preposition meaning that can be used to describe a compound expression through a gamification mechanism,
- the representation of taxonomic as well as transversal (non-taxonomic) relations extracted by using FCA and RCA techniques not previously applied in Arabic studies.

3. Background

3.1. Background on Ontologies

In Artificial Intelligence, an ontology is, according to Tom Gruber, "the specification of conceptualizations used to help programs and humans share knowledge" (Gruber, 1995). According to this definition, an ontology is a set of precisely-specified concepts and relations employed in order to create an agreed-upon vocabulary and semantic structure for exchanging information about this domain.

A concept represents a set or class of entities within a domain. Relations describing the interactions between concepts can be classified into two major categories: taxonomic relations and non-taxonomic ones. The former organize concepts into hierarchical tree structures, such as specialization relations commonly known as the "is a/ kind of" relations. However, the latter relate concepts across tree structures like locative (space and time) and causative relations.

Ontologies play an increasingly pervasive role in the modern knowledge-based systems as they constitute a powerful tool for supporting natural language processing, information retrieval and text mining.

Ontology learning has been widely studied in literature so that a diverse spectrum of approaches classifications were developed based on different criteria, such as the degree of automation and the types of input knowledge resources. According to the last criterion, several studies were performed on unstructured data, like text documents, semi-structured data, as Web pages and dictionaries, as well as on structured data such as object-oriented data or knowledge models (Kumova, 2015).

Ontology learning from unstructured texts is the most prevalent process thanks to the accessibility and availability of texts in different domains. In addition, texts present a good carrier of stabilized and shared knowledge. We illustrated the process of ontology learning from texts that are usually decomposed following different steps and based on the ontological element learned in Mezghanni and Gargouri (2014).

3.2. Background on FCA/RCA

FCA is a mathematical theory used to identify all the possible groupings having common properties. As revealed in Fig. 1, the main notions of the FCA are the formal context, formal concept and concept lattice.

A triplet = O, A, R is a (formal) context if

- O is a set of objects;
- A is a set of attributes and
- R is a binary relation (OxA) called "Incidence"

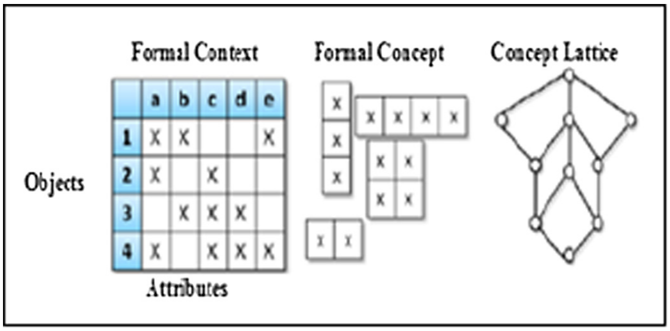
The formal context is usually described by a cross-table (or incidence matrix) where rows and columns represent respectively the objects and the attributes of the context. A cross, in column m of row g, means that R = "object o has attribute a" or "attribute a is true for object o". The absence of a cross means that R = "o does not have attribute a". An example of such table representing a formal context of 5 objects described by 4 attributes is shown in Table 1.

Thus, in this table, object o5 has attribute a1, but it does not have attribute a2.

From the formal context, we calculate the formal concept described as a pair C = E, I where E is the maximal collection of objects (called the extent) sharing common attributes I (called the intent).

The whole set of red cells, in Table 2, represents formal concept C6(E1, I1) = (o1, o2, o3, o4, a3, a4). It is worth-noting that there are further formal concepts.

[4] http://www.nooj4nlp.net.



	a	b	c	d	e
1	X	X			X
2	X		X		
3		X	X	X	
4	X		X	X	X

Fig. 1. Foundational elements of FCA.

Table 1
Cross-table describing formal context.

R	a1	a2	a3	a4
o1	X	X	X	X
o2	X		X	X
o3		X	X	X
o4		X	X	X
o5	X			

Table 2
A formal concept C6.

R	a1	a2	a3	a4
o1	X	X	X	X
o2	X		X	X
o3		X	X	X
o4		X	X	X
o5	X			

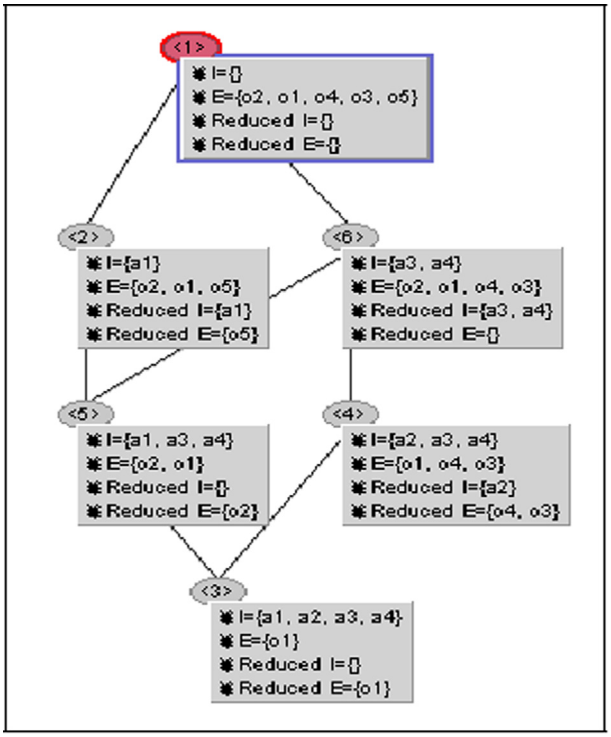


Fig. 2. Concept lattice L(C_K).

From the set of all concepts, we can derive a conceptual hierarchy in a complete lattice structure called concepts lattice or Galois lattice L(C_K)to illustrate the hierarchy relation among the groups (The extents and intents) (Bouhriz et al., 2015). The corresponding concept lattice L(C_K), designed using Galicia platform,[5] is depicted in Fig. 2 where we can see the concept C6 described above.

RCA (Huchard et al., 2007), an extension of FCA, is an original approach applied to extract formal concepts from sets of data described by binary and relational attributes, to model these links and then to infer relations between formal concepts whose semantics is similar to roles in ontologies.

A RCA is represented through a Relational Context Family (RCF) which involves a collection of contexts describing the entities of different categories and the relations between them.

As formally described by Mezghanni and Gargouri (2014), a RCF is a pair (K, R) where K is a set of formal (object-attribute) contexts $K_i = (O_i, A_i, I_i)$ and R is a set of relational (object-object) contexts $r_{ij} \subseteq O_i$ x O_j, where O_i(domain of r_{ij}) and O_j(range of r_{ij}) are the object sets of the contexts K_i and K_j, respectively.

RCF is used in an iterative process to produce, at each step, a set of concept lattices. First, the building concept lattices is entirely based on the formal contexts. In the subsequent steps based on a scaling mechanism, all the links between the objects are transformed into conventional FCA attributes. A collection of lattices, whose concepts are linked by relations, is derived. The steps are reiterated until the stability of lattices is achieved when no more new concepts are generated. More details on RCA process are presented in Section 5.

4. Approach

Our strategy follows a common statement suggesting that some linguistic constructs reliably convey the same type of knowledge, such as semantic or ontological relations (Aguado de Cea et al., 2009). Indeed, the main idea behind our approach is to exploit internal structures of compounds considered as the most meaningful entities for deciphering semantic relations. Regarding the importance of the Arabic language peculiarities, summarized in Mezghanni and Gargouri (2016), and the possibility of using linguistic knowledge acquired for one natural language processing task (which is in our case concept extraction), we adopted the rule-based approach. It relies on a core of solid linguistic knowledge which usually provides highly accurate results. After that, FCA/RCA approaches, discussed in Section 5, are applied.

4.1. General relations deciphering

In our research, we focus essentially on Arabic compound nouns (2-gram up to 5-gram) having the internal structure presented in Table 3. These structures are generated through a set of POS patterns using the platform NooJ. We distinguish different types of compounds: adjective (مركب نعتي), prepositional (مركب حرفي), annexation (مركب إضافي), etc.

In this table, N stands for Noun, ADJ denotes Adjective, ADV represents Adverb, PREP refers to Preposition and PREF corresponds to the definite article (ال/al). The examples are provided with their english translation and their transliteration using Xerox Morphology System.[6]

A compound noun contains normally two parts. In Arabic language, the first part is indispensable as it represents the head identifying an object or a person. However, the second part modifies or describes the object or person in question. According to the syntactic category of the second part, we constructed a set of 12 rules

[5] http://www.iro.umontreal.ca/g̃alicia/.

[6] https://open.xerox.com/Services/arabic-morphology/Consume/ Morphological%20Analysis-218

generating numerous relations. It should be noted that the rules of Arabic grammar are behind these inferences.

We recognize two types of semantic relations: specific and general. The former are hierarchical and represented by the subClassOf relation (is-a relation) representing the relation between the compound and its second element. Example: A subClassOf B where A and B refer respectively to the specific entity type and the generic entity type (i.e., Manager subClassOf Employee). However, the latter is transversal and denoted by an objectProperty relation representing the possible semantic links between the elements of the compound. For example: A objectProperty B states that there is a relation between A and B.

```
Rule 1:
IF Compound [N+ADJ] THEN
{ subClassOf(Compound, N);
// Cons: N is a concept }
```

```
Rule 2:
IF Compound [N1+N2] THEN
{ objectProperty(N1, N2);
// Cons: N2 is a concept
subClassOf(Compound, N1);
// Cons: N1 is a concept }
```

In the following part, we will show that Compound [Pattern] designates a compound composed of the elements of the pattern and Cons indicates that a Constraint is imposed.

Rule 1 states that there is a subClassOf relation between the compound and the noun N of the compound. This relation is deduced by defining the adjective noun which introduces specialization relation between the noun (المنعوت) and the adjective (النعت) since the adjective noun brings out from a general noun to a more specific one. For example, from the compound دعوى عمومية [دعوى+عمومية], we derive the relation: subClassOf(دعوى, دعوى عمومية) which designates in English that public action is a subClassOf action.

We added a restriction to the rule to check if the N presents a concept from the domain corpus. Otherwise, this derivation will not be interesting. In the above example, we have to check if action is a concept in our corpus. In this case, action (دعوى) is a domain concept and it designates "a judicial proceeding brought by one party against another". In the legal terminology, we distinguish between civil action (دعوى مدنية) and public action (دعوى عمومية). This distinction approves the above-described rule.

Rule 2 states that, in a noun-noun compound, there is a subClassOf relation between the first noun and the compound. This relation is motivated by defining the determinative compounds which introduce hyponymy between the compound and its second noun. For example, from the compound حاكم التحقيق, we derive the relation: subClassOf(حاكم,حاكم التحقيق), whose translation into English means that investigating judge is a subClassOf a judge.

```
Rule 3:
IF Compound [N1+ADJ+N2] THEN
{ objectProperty(Compound [N1+ADJ], N2);
// Cons: N2 is a concept
subClassOf(Compound, Compound [N1+ADJ]);
Rule1 (Compound [N1+ADJ]); }
```

```
Rule 4:
IF Compound [N1+N2+ADJ] THEN
{ subClassOf(Compound, N1);
// Cons: N1 is a concept
objectProperty(Compound [N2+ADJ], N1);
Rule1 (Compound [N2+ADJ]); }
```

```
Rule 5:
IF Compound [N1+N2+N3] THEN
{ subClassOf(Compound, N1);
// Cons: N1 is a concept
objectProperty(Compound [N2+N3], N1);
Rule1 (Compound [N2+N3]); }
```

```
Rule 6:
IF Compound [N1+ADJ1+ADJ2]
THEN
{ subClassOf(Compound, Compound [N1+ADJ1]);
Rule1 (Compound [N1+ADJ1]);
}
```

Table 3
Different POS Patterns used to extract compounds, with occurrence examples.

Pattern of the compound	Example / Transliteration	English translation
2-gram		
N+ADJ	ضابطة عدلية/ DAbTp Edlyp	Judicial Police
N+N	رئيس الجمهورية/ rgys Aljmhwryp	President of Republic
3-gram		
N+ADJ+N	مدير عام السجون / mdyr EAm Alsjwn	Director General of Prisons
N+PREP+N	قرار بالحفظ / qrAr bAlHfZ	Decision of conservation
N+ADJ+ADJ	محكمة مدنية مختصة / mHkmp mdnyp mxtSp	Competent civil court
N+N+ADJ	مأمورالضابطة العدلية / mmwr AlDAbTp AlEdlyp	Judicial Police Officer
N+N+N	رئيس مركز الشرطة / rgysmrkzAl$rTp	Police chief
4-gram		
N+N+N+ADJ	رئيس مركز الحرس الوطني / rgys mrkz AlHrs AlwTny	Chief of the National Guard
N+N+ADJ+ADJ	عضوية الهيئات الدستورية المستقلة / EDwyp AlhygAt Aldstwryp Almstqlp	Adhesion of independent constitutionals organs
N+N+N+N	عضوية مجلس نواب الشعب / EDwyp mjls nwAb Al$Eb	Adhesion of people's congress
N+N+N+ADJ	آجال سقوط الدعوى العمومية / —jAl sqwT AldEwY AlEmwmyp	Prescription of pubic action
N+ADJ+N +ADJ	الرائد الرسمي للجمهورية التونسية / AlrA}d Alrsmy lljmhwryp Altwnsyp	Official Journal of the Republic of Tunisia
5-gram		
N+N+PREP +N+ADJ	كاتب الدولة للشؤون الخارجية / kAtbAldwlpll$&wnAlxArjyp	Secretary of State for Foreign Affairs

On the other hand, this compound expresses an additional relation between the two involved nouns. But, this relation is not linguistically explicit. Applying this rule on the same compound حاكم التحقيق (investigating judge), we will have a relation between حاكم التحقيق (investigating judge) and التحقيق (investigating). The example expresses a possession relation. However, we cannot consider this case as a general one. Indeed, noun-noun compound do not involve only possession relation as in girl's dress, but also other relations such as containment relation as in girl's face and location. For example, from محكمة الإستئناف (Appeal court) we cannot say appeal possess court; whereas we can understand that appeal is carried out in court.

All other rules concerning n-gram compounds with n> 2 are based on the rules for (n-1)-gram.

For instance, the 3-gram Compound [N1 + N2 + ADJ], presented in Rule 4, can be viewed as Compound [N1 + N2] if we consider (N2 + ADJ) as one entity N2.

That is to say:مأمور| الضابطة العدلية(judicial police – officier)

N2 – N1 → Rule 2.

With N2 = N + ADJ → Rule 1.

For the Compound [N1 + PREP + N2], illustrated in Rule 7, we focus on the most common prepositions in our corpus:"في" ,"ب", "ل", "من".

Besides, the 4-gram Compound [N1 + N2 + N3 + N4], presented in Rule 10, can be viewed as Compound [N1 + N2] if we consider (N1 + N2 + N3) as one entity N2.

That is to say:عضوية | مجلس نواب الشعب(adhesion – of people's congress)

N2 – N1 → Rule 2.

With N2 = N + N + N → Rule 5.

```
Rule 8:
IF Compound [N1+N2+N3+ADJ] THEN
{ subClassOf(Compound, N1);
// Cons: N1 is a concept
objectProperty(Compound [N2+N3+ADJ], N1);
Rule6 (Compound [N2+N3+ADJ]);
}
```

```
Rule 9:
IF Compound [N1+N2+ADJ1+ADJ2] THEN
{ subClassOf(Compound, N1);
// Cons: N1 is a concept
objectProperty(Compound [N2+ADJ1+ADJ2], N1);
Rule5 (Compound [N2+ADJ1+ADJ2]);
}
```

```
Rule 7:
IF Compound [N1+PREP+N2] THEN
{ subClassOf(Compound, N1);
// Cons: N1 is a concept
Switch(PREP) {
case ("في"): R1|R2|R3|R4|R5 (N1, N2); break;
case ("ب"): R6|R7|R8 (N1, N2); break;
case ("ل"): R9 (N1, N2); break;
case ("من"): R10|R11|R12 (N1, N2); break;}
// Cons: N2 is a concept }
```

```
Rule 10:
IF Compound [N1+N2+N3+N4] THEN
{ subClassOf(Compound, N1);
// Cons: N1 is a concept
objectProperty(Compound [N2+N3+N4], N1);
Rule7 (Compound [N2+N3+N4]); }
```

Additionally, the 5-gram Compound [N1 + N2 + PREP + N3 + ADJ], showed in Rule 11, can be viewed as Compound [N1 + PREP + N2] if we consider (N1 + N2) as one entity N2 and (N3 + ADJ) as one entity N1.

That is to say:كاتب الدولةال | لشؤون الخارجيّة(adhesion – of people's congress)

N2 – PREP – N1 → Rule 7.

With N2 = N + ADJ → Rule 1.

```
Rule 11:
IF Compound [N1+N2+PREP+N3+ADJ] THEN
{ subClassOf(Compound, N1);
// Cons: N1 is a concept
objectProperty(Compound [N2+N3+N4], N1);
Rule2 (Compound [N1+N2]);
Rule1 (Compound [N3+ADJ]); }
```

```
Rule 12:
IF Compound [N1+ADJ1+PREP+N2+ADJ2] THEN
{ subClassOf(Compound, N1);
// Cons: N1 is a concept
objectProperty(Compound [N1+ADJ], Compound [N2+ADJ2]);
Rule1 (Compound [N1+ADJ]);
Rule1(Compound [N2+ADJ2]); }
```

To generalize these rules, we implemented a Java program that accepts a list of generated compounds as input and a list of the resulting relations as output. For each compound, the program applies the corresponding rules. Each deduced relation is verified. If it is not deduced, it will automatically be added to the list of the resulting relations. Otherwise, it will be deleted.

4.2. General objectProperty relations specification

Obviously, the derived objectProperty relations are very general. The simple existence of such relation is not enough and can lead to ambiguity. In order to adequately precise them, we adopted a data-driven strategy to find the preposition that can most likely be used to specify "objectProperty (X, Y)" expression. The semantic of the chosen preposition is considered as the specific relation.

By definition, a preposition is a word or set of words that usually precedes, a noun or pronoun and expresses a relation with another word or element in the clause (Litkowski, 2002). Thus, the preposition expresses many semantic relations between the constituents that it relates.

The addressed task is to specify adequately the general recognized relations between the nouns that form the compound nouns. We performed this task using the relations expressed by prepositions summarized in Table 4.

In this context, we developed a simple user-interface called "GUESS" involving thirty non-linguisitic players. The players are asked to vote through selecting the most appropriate prepositions that may join the constituents of compounds.

To make easier the vote, we gave examples for each preposition use, which also helped us to be very specific and precise as shown in Fig. 3. Indeed, no selection implies that no choice can express the meaning of the relation.

We applied this strategy since it is extremely easy, straightforward and useful, especially without linguistic training. Once the judgments are collected, we obtain a binary matrix with the compounds in the rows, R1 to R12 in the columns. Based on this matrix, the agreement score between players, defined as the number of the similar judgments, is computed. For this reason, the pair occurring more frequently is considered. When ties occur, one of the pairs will be arbitrarily chosen. Finally, the "objectPropety" is replaced by the corresponding relation.

5. Experimentation results

As we previously indicated, this work is part of the process of ontology construction from texts. The corpus consists of 50 articles from the Criminal Procedures Penal Code and 20 Criminal Law Decisions of the cassation Court gathered from the official Tunisian portal.[7] We consider, in this paper, the case of legislative documents "code" which contains a division into six gradations: Code, Book, Chapter, Section, Sub-section, and Article. We also take into account the article as the leaf tags (a node with no children).

To deeply explain the experiment, the subsequent steps are schematized in Fig. 4.

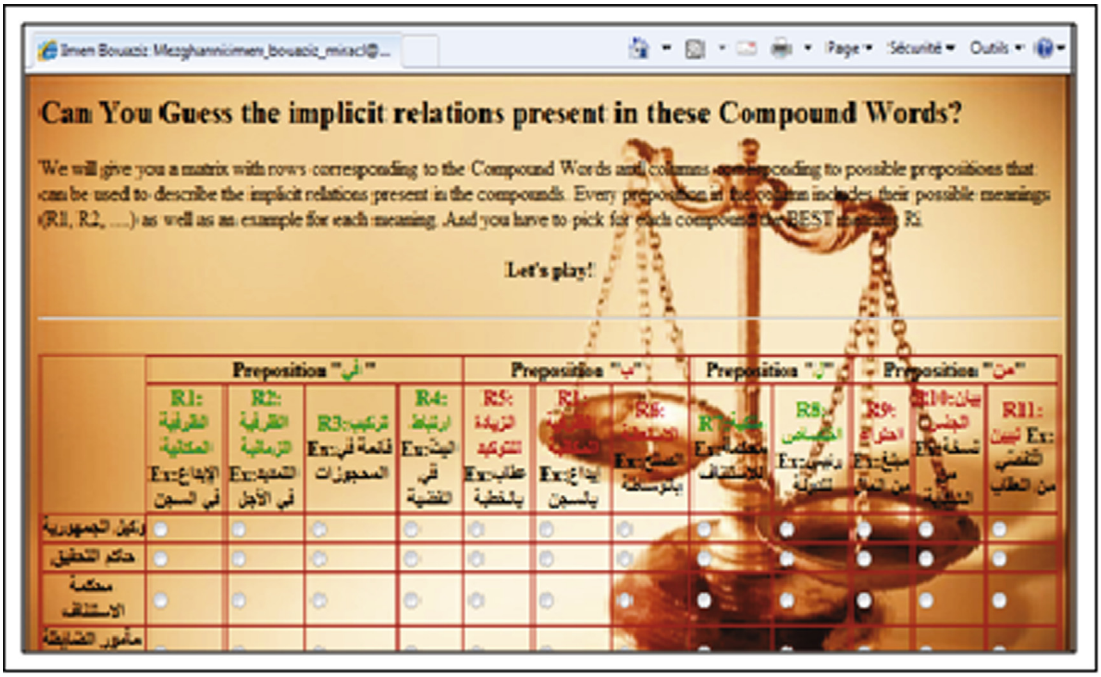
Therefore, we focus on the two following most important points in this experiment:

- the deduction of relations using a mechanism of human-vote GUESS,
- the implementation of FCA/RCA techniques based on the results provided by considering the first point.

Table 4
Prepositions Semantics.

Relation designation	Example	English translation	Ref
The preposition "في":			
Spatial relation الظرفية المكانية	إيداع في السجن / <ydAE fy Alsjn	Committal to prison	R1
One object located or surrounded in space by another.			
Temporal relation الظرفية الزمانية	تمديد في الأجل / tmdyd fy Al>jl	Extension of the deadline	R2
One object occurred in time by another.			
Composition relation علاقة تركيب	قائمة في المحجوزات / qA}mp fy AlmHjwzAt	List of seized items	R3
One object assembling several components to form a unified whole.			
Association relation علاقة ارتباط	بتّ في القضية / btf̃y AlqDyp	Disposition of the case	R4
One object connected or attached to another.			
The preposition "ب":			
Increase emphasis الزيادة للتوكيد	عقاب بالخطية / EqAb bAlxTyp	Imposition of a fine	R5
One object used to give focus to another.			
Spatial relation الظرفية المكانية	إيداع بالسجن / <ydAE bAlsjn	Committal to prison	R6
Usage relation الاستعانة	صلح بالوساطة / SlH bAlwsATp	Conciliation through mediation	R7
One object indicating the use of another.			
The preposition "ل":			
Possession relation علاقة ملكية	محكمة الإستئناف / mHkmp AlAst}nAf	Court of Appeal	R8
One object owns another.			
Competence relation علاقة اختصاص	رئيس للدولة / r}ys lldwlp	President of the Republic	R9
Object indicating Skills in a specialized area or profession.			
The preposition "من":			
Containment relation علاقة إحتواء	مبلغ من المال / mblg mn AlmAl	Sum of money	R10
Object used to indicate an amount or number of another.			
Typology relation بيان الجنس	نسخة من الشكاية / mblg mn AlmAl	Copy of complaint	R11
Object used to indicate the type or kind of another.			
Indication relation علاقة تبيين	تفصّي من العقاب / mblg mn AlmAl	Escape from punishment	R12

[7] www.e-justice.tn



Can You Guess the implicit relations present in these Compound Words?

We will give you a matrix with rows corresponding to the Compound Words and columns corresponding to possible prepositions that can be used to describe the implicit relations present in the compounds. Every preposition in the column includes their possible meanings (R1, R2, ...) as well as an example for each meaning. And you have to pick for each compound the BEST meaning Ri.

Let's play!

	Preposition "في"				Preposition "ب"			Preposition "ل"		Preposition "من"		
	R1: الظرفية المكانية Ex: الإيداع في السجن	R2: الظرفية الزمانية Ex: التمديد في الأجل	R3: تركيب Ex: قائمة في المحجوزات	R4: ارتباط Ex: البت في القضية	R5: الزيادة التوكيد Ex: عقاب بالخطية	R1: الظرفية المكانية Ex: إيداع بالسجن	R6: الاستعانة Ex: الصلح بالوساطة	R7: تعليل Ex: مطلب للاستئناف	R8: الاختصاص Ex: رئيس الدولة	R9: احتراز Ex: مبلغ من المال	R10: بيان الجنس Ex: نسخة من الوثيقة	R11: تبيين Ex: التفصي من العقاب
وكيل الجمهورية	○	○	○	○	○	○	○	○	○	○	○	○
حاكم التحقيق	○	○	○	○	○	○	○	○	○	○	○	○
محكمة الاستئناف	○	○	○	○	○	○	○	○	○	○	○	○
مأمور الضابطة	○	○	○	○	○	○	○	○	○	○	○	○

Fig. 3. GUESS interface.

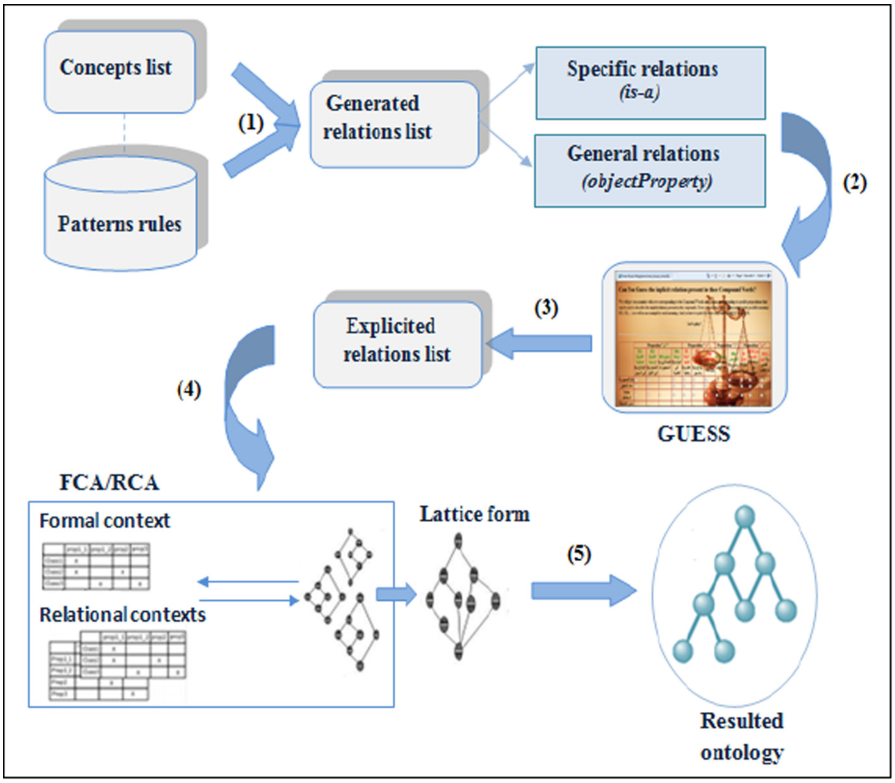


Fig. 4. Process of semantic relations extraction.

5.1. GUESS interpretations

The number of the different compounds on which we experimented the method is 56. Thirty players had to choose one meaning from 12 different semantics of the prepositions illustrated with examples. The distribution of R_i is presented in Table 5 where the second column displays the total number of times and given R_i is hand-picked by players.

The analysis of the results divulges different interpretations. Unexpectedly, all the prepositions were picked. On the one hand, there were 41 different compounds (73%) that had unanimous vote among various players. This can be mainly explained by the

Table 5
Choice of $R_i$through GUESS.

R_i	Picked
R_1	46
R_2	10
R_3	65
R_4	77
R_5	30
R_6	46
R_7	12
R_8	298
R_9	51
R_{10}	07
R_{11}	33
R_{12}	11

highest score for the common preposition selected ل with the possession meaning as it is the case in: حاكم للتحقيق (judge of investigating), وكيل للجمهورية (attorney of republic), مركز للشرطة (station of police).

This broad convergence of views shows the advantage of using prepositions semantics to uncover relations and the effect of specifying the given choices in GUESS. We think this agreement is encouraging. On the other hand, there were some compounds that got more than one preposition. It seems accurate that إيداع في السجن or إيداع بالسجن (committal to prison) both expressed a spatial relation, and both prepositions would probably be correct. Moreover, it is apparent that not all prepositions were used with equal frequency. The least common picked preposition was من.

All the answers were collocated and then introduced to two expert evaluators in linguistic and legal fields. The role of the linguist and the lawyer was then to validate and exactly to verify the accuracy of the chosen relations. The collaboration between linguist and domain experts ensured both linguistic quality and reliability of technical information. The experts independently assigned a binary value accordingly to their agreement on the result of GUESS. The agreement reached 84%, confirming the performance of the pursued strategy. At this stage, we simply gathered all the non-contentious cases and we did not decide how to solve the rest of the cases.

5.2. Implementation of FCA

To make the results more readable and interpretable, we took into account just a limited number of concepts with the list of the generated relations already validated by experts. However, it must be accentuated that this structure is partial. Thus, it must also consider other types of relations.

The concepts were initially modeled as a formal context. The objects correspond to the concepts, while attributes are the structural positions (articles tags) where the concepts appear, which corresponds to the actual relations of the corpus of legal documents processed in this work.

The formalization of the concepts is given by the formal context $K_{Concepts}$ = (O, A, R), where O is a set of concepts (دعوى عمومية (public action), حاكم التحقيق (investigating judge), etc.), A is a set of articles tags of the documents (e.g. Article 1, Article 2, etc.). It means that the concept o is structurally characterized by its presence in the article a.

Table 6 illustrates the formal context Concepts. In this table, due to lack of space, we denote by the numbers in columns the Articles numbers. Besides, we replace the concepts by alphabetic letters.

Table 6
Formal context of Concepts.

Concepts	7	9	10	11	16	36	48	104	277
a (مساعد)								X	X
b (محكمة)									X
c (مأمور)					X				
d (رئيس)							X		
e (تحقيق)						X	X		
f (محكمة مدنية)	X								
g (ضابطة عدلية)		X							
h (وكيل الجمهورية)			X	X					
i (مركز الشرطة)			X	X					
j (محكمة مدنية مختصة)	X								
k (رئيس مركز الشرطة)				X					
l (مأمور الضابطة العدلية)			X	X					
m (وكيل عام للجمهورية)			X	X					
n (مساعد وكيل الجمهورية)			X	X					

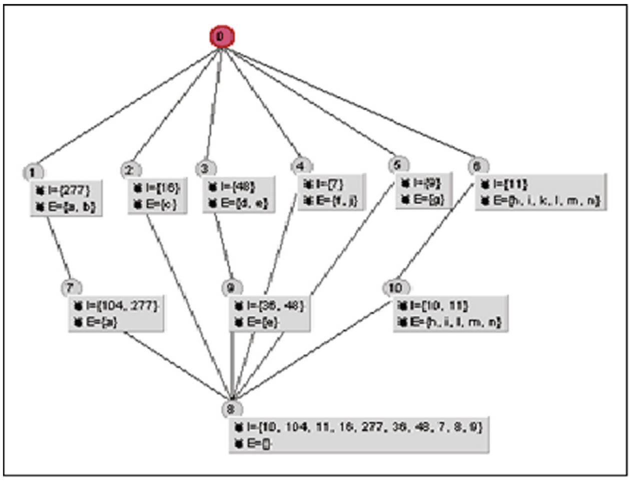


Fig. 5. Lattice L ($C_{KConcepts}$, $\leqslant K_{Concepts}$).

Fig. 5 illustrates the concept lattice L ($C_{KConcepts}$, $\leqslant K_{Concepts}$) corresponding to the formal context of the concepts $K_{Concepts}$given by Table 6. This lattice is represented by a Hasse diagram in which nodes are the concepts and edges are the links of specialization/-generalization through Galicia platform.

5.3. Implementation of RCA

As mentioned formerly, the relations modeled through RCA are not only (objects x objects), but also (attributes x attributes). We are currently studying the first kind of relations which expresses the links between our multi-word terms concepts. The second kind handles the cross-reference between articles which is highly propagated in the legal codes, which may be studied in the future works.

From a collection of contexts and of inter-context relations, RCA builds a RCF. This family is the starting point of the process of constituting the relational lattice families whose concepts are linked by relations.

In the above-illustrated example, diverse relations are generated from the compounds. We present, in Table 7, an example of these relations علاقة ملكية (has-a) relating numerous concepts. However, in Table 8, we present the relation جنس من (subClassOf). The instances of this relation are called "links". Table 7 shows the علاقة ملكية (مأمور (c), ضابطة عدلية (g)) which is generated from the compound مأمور الضابطة العدلية (l).

Table 7
Inter-context relation علاقة ملكية (has-a).

علاقة ملكية	a	b	c	d	e	f	g	h	i	j	k	l	m	n
a														
b														
c														
d														
e														
f														
g			X											
h	X													
i				X										
j														
k														
l														
m														
n														

Table 8
Inter-context relation جنس من (subClassOf).

جنس من	a	b	c	d	e	f	g	h	i	j	k	l	m	n
a														
b														
c														
d														
e														
f		X												
g														
h														
i														
j						X								
k				X										
l			X											
m														
n	X													

Both relations together with the context Concept form our RCF sample:

- Context: $K_{Concepts}$(Concepts X Articles).
- Relations:

  علاقة ملكية (has-a) $\subseteq ConceptsXConcepts$

  جنس من (subClassOf) $\subseteq ConceptsXConcepts$

The main steps involved in RCA are based on multi-FCA method, producing thus a set of lattices called Concept Lattice Family (CLF). Fig. 6 outlines the Multi-FCA method (Falih and Omar, 2015) that describes the step-wise construction of the fix point solution from the initial RCF. The iterative logic of this technique generates, at each step, a set of concept lattices.

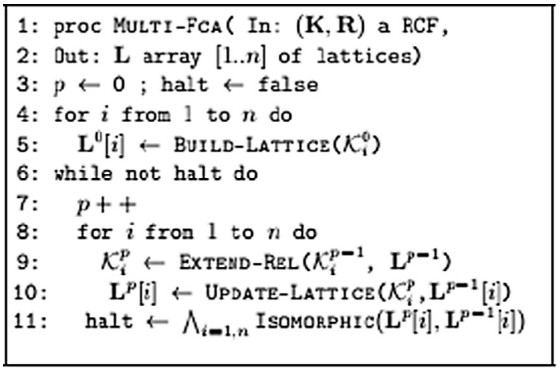
It is described by Dolques et al. (2013) as follows:

Step 0:

– Apply FCA on the contexts from K to build lattices.

Step > 0:

– Extend each formal context with the lattices from previous step and relational contexts through a relational scaling mechanism used to translate links into conventional context attributes;
– Apply FCA on each extended context to get new lattices whose concepts are linked by relations;
– Stop when a fix-point is obtained: lattices are isomorph between two consecutive steps and leaves unchanged concept extents.



```
proc MULTI-FCA( In: $(\mathbf{K},\mathbf{R})$ a RCF,
Out: $\mathbf{L}$ array $[1..n]$ of lattices)
$p \leftarrow 0$ ; halt $\leftarrow$ false
for $i$ from 1 to $n$ do
    $\mathbf{L}^0[i] \leftarrow$ BUILD-LATTICE($\mathcal{K}_i^0$)
while not halt do
    $p++$
    for $i$ fron 1 to $n$ do
        $\mathcal{K}_i^p \leftarrow$ EXTEND-REL($\mathcal{K}_i^{p-1}$, $\mathbf{L}^{p-1}$)
        $\mathbf{L}^p[i] \leftarrow$ UPDATE-LATTICE($\mathcal{K}_i^p$,$\mathbf{L}^{p-1}[i]$)
    halt $\leftarrow \bigwedge_{i=1,n}$ ISOMORPHIC($\mathbf{L}^p[i]$, $\mathbf{L}^{p-1}[i]$)
```

Fig. 6. The RCA process (Huchard et al., 2011).

Therefore, in our example, the first step consists in building the concept lattice according to the principles of the FCA from binary as shown in Table 6.

While scaling the علاقة ملكية (has-a) relation, the object g is linked to c belonging to the extents of concepts C0 and C2 in the initial lattice presented in Fig. 5, h is linked to a belonging to the extents of concepts C0, C1 and C7 while i is linked to d belonging to the extents of concepts C0 and C3. Thus, relational information is incorporated into the scaled version of the contexts. For instance, the relational attributes.

- علاقة ملكية: C0 and علاقة ملكية: C2 are assigned to the object g (C5 in the final lattice),
- علاقة ملكية: C0, and علاقة ملكية: C1 and علاقة ملكية: C7 are assigned to the object h (C13 in the final lattice),
- علاقة ملكية: C0 and علاقة ملكية: C3 are assigned to the object i (C14 in the final lattice).

This incorporation leads to additional attributes shared among objects and hence new concepts emerge. By factoring out the new attributes into concept intents, object links are lifted up to the concept level, which yields relations between concepts and justifies the extension shown in the final lattice illustrated in Fig. 7.

The same process was applied to the جنس من (subClassOf) relation where f is linked to b, j to f, k to d, l to c and n to a, which leads to the creation of new attributes جنس من: C0, جنس من: C1, جنس من: C2, جنس من: C3, جنس من: C4, جنس من: C7.

The concept C10 represents the concepts h, i, l, m and n which belong to the same Articles 10 and 11. In the final lattice:

- C10 is categorized into C11 and C15 to consider the two relations علاقة ملكية (has-a) and جنس من (subClassOf).
- C11 is enriched by the relational attribute علاقة ملكية: C0, which basically means that these concepts are also in relation علاقة ملكية with others.
- C15 is enriched by the relational attribute جنس من: C0, which essentially means that these concepts are also in relation جنس من with others.

6. Evaluation and discussion

Unlike the works proposed in the literature (Al-Yahya et al., 2016; Al Zamil and Al-Radaideh, 2014) to evaluate our approach, we conducted a translation of RCA constructs to the corresponding ontological components. The target ontology was compared to a hand-crafted ontology, manually created by researchers in the ontology field within our laboratory.

6.1. Ontology derivation

The final lattice can be considered as the knowledge model from which we can build the ontology (Bendaoud et al., 2008; Bendaoud et al., 2007).

To represent the formal concepts of the lattice, we have to choose a Knowledge representation language based on description logics (DL) formalism.

The considered target DL is $\mathcal{LFE}$. This formalism includes constructors T (top), ⊥ (bottom), C ∩ D (conjunction of concepts) and ∀ r.C∃ r.C (universal and existential quantifiers). This minimum

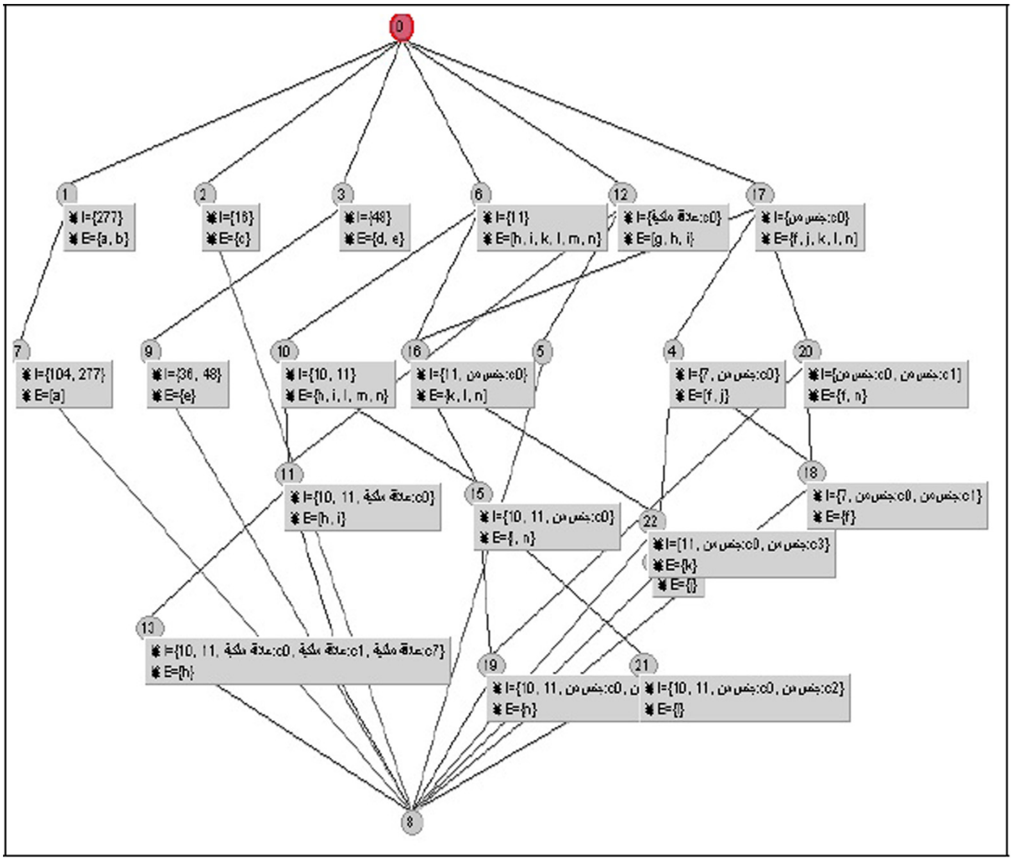


Fig. 7. The lattice obtained by applying RCA.

set of constructors is large enough to represent all elements from the final concept lattice.

Following the strategy of Huchard et al. (2003), the translation between the RCA elements and the $\mathcal{LFE}$ was carried on using a function α defined as follows:

α: (K, R) → TBox∪ABox, where: (K, R) is a family RCF, TBox (conceptual expressions) and ABox (the set of ground facts) are the components of the ontology.

Our "α" transformation works differently as follows:

– Each context is translated into an atomic concept that expresses the top T of the hierarchy in this context c ≡ α(K),
For example: α (Concepts)≡ Concepts.
– Each formal attribute is translated into an instance α(m) in the ABox.
For instance, the attribute Article 10 becomes the instance α (Article 10).
– Each relation r ∈R is translated into a primitive role α(r)).
For example, علاقة ملكية has a primitive role in the TBox.
– Each relational attribute r.C is transformed in the TBox into an atomic role with universal or existential quantification, depending on the scaling schema for inter-context relations c ≡ α(r) ≡ ∃ r.α(c).
For instance, α(علاقة ملكية.d) ≡ ∃ علاقة ملكية.d
– Each formal concept C = (E, I) ∈C is transformed in the TBox into a defined concept formed by the conjunction of primitive concepts and existential role quantications α(c) ≡ $\cap_{m \in I}$ C_α(m).
For example, α(C7) ≡ ∃ C_277.T∩ ∃ C_104.T
–Each subsumption relation between concepts C1 ⊆ C2 is transformed into an Inclusion axiom α(C1) ⊆ α(C2).
For instance, α(C7) ⊆ α(C1).
–Each object g ∈G is transformed into a defined concept c ≡ α(g) ≡ ∃ g.T.
For example, α(16) ≡ ∃ C_16.T.

– All the formal attributes, translated through the function α into instances, are connected to a general defined concept. In our case, we initiated it "Articles".
– Each atomic concept, representing a context, is connected to the general concept representing the translation of the formal attributes (in our case Articles) via a descriptive relation.

The application of the function α to the lattice of Fig. 7 produced the ALO (Arabic Legal Ontology) ontology to which we added the following rules:

In order to visualize the created ontology, we used the Protégé ontology editor[8] with its jambalaya plug-in, which supports the visualization of Arabic letters. Fig. 8 shows only a small portion of the produced ontology. The whole ontology contains 92 concepts and 145 relations. As illustrated in Fig. 8, the ontology contains hierarchies between concepts. For example, C_10 is a specific type of C_11. Thus, it inherits the properties of the C_11 concept. C_10 is called a child concept, C_11 is named a parent concept, and their relationship is captured by an is-a arrow (inheritance). Additionally, since C_10 and C_11_جنس constitute together the collection of C_11, C_11 is a union concept and C_10, C_11_جنس are the corresponding member concepts. This relationship is represented by the unionOf arrows (membership). Such inheritance and membership relationships are frequently encountered in real-life ontologies. The concepts were labeled to help experts read the ontology.

6.2. Ontology evaluation

Despite the fact that approaches dealing with the evaluation of ontologies are numerous, no standard methodology has been agreed upon. After ontology derivation, we need to assess how good the generated ALO ontology reflects the legal domain. To achieve this goal, we compared a part of ALO and a hand-crafted corresponding ontology called CrimAr (Criminal Arabic Ontology) manually created by a lawyer and a group of researchers in our laboratory specialized in ontology. The CrimAr ontology, shown in Fig. 9 in which we highlighted the same concepts of our example, was considered as a reference through which we can judge and evaluate the performance of the derived ontology. Usually, the evaluation of ontology comparison relies on the precision and recall as the most well-known measures originating from information retrieval.

This pair of variables measures is based on the comparison of an expected result and the effective result of the evaluated system, with precision being the proportion of the retrieved documents that are relevant and recall being the proportion of the relevant documents that have been retrieved. In logical terms, precision is supposed to measure the correctness of the evaluated system, while recall is supposed to quantify its completeness. Since these measures are commonly used and well understood, they have been adhered and adapted for ontology comparison evaluation (Do et al., 2003). In this paper, we focus only on the ontology evaluation at *the relational level*. Thus, the precision of a given relation was measured as the percentage of correct discovered relations over the total number of discovered relations, while recall was measured as the percentage of correct discovered relations over the total number of relations of the reference ontology.

Our experiments on the whole ontology shows a large number of conceptual relations. The herein presented work evaluates the success of only the relations of the above-described example which are علاقة ملكية (has-a) and علاقة جنس من (subClassOf). Their corresponding evaluation results are reported in Table 9.

As expected, the obtained evaluations of experimental results are very encouraging. The evaluation highlights the importance of the produced ontology in terms of relations. The recall achieved by جنس من (subClassOf) outperforms that of علاقة ملكية (has-a). Evenly, the precision, reached by جنس من (subClassOf) exceeds that of علاقة ملكية (has-a). The good results obtained by applying the proposed method on this relation are explained by the fact that this relation is obtained from the majority of the studied compounds.

Compared to the manually created ontology, it can be noticed that practically the same concepts exist but the relations are different. CrimAr ontology relates the majority of concepts by hierarchical relations to indicate supervision authorities, according to the competencies for the magistrates such as مأمور الضابطة العدلية (judicial police officer),للجمهورية وكيل عام (prosecutor general of the republic), مساعد وكيل الجمهورية (Assistant Attorney General);

[8] http://opprotege.stanford.edu.

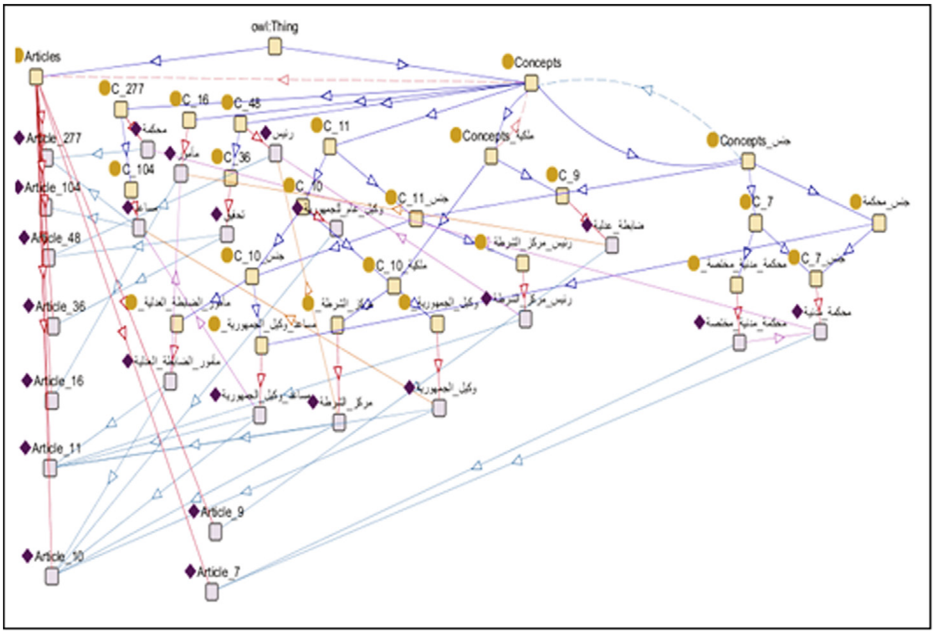


Fig. 8. The generated ontology.

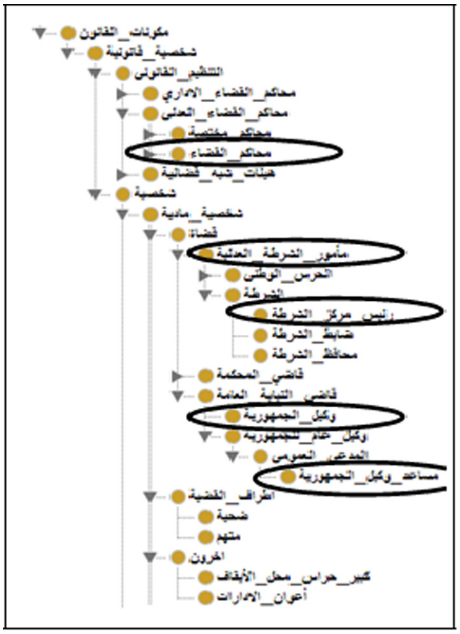


Fig. 9. Subset of the CrimAr Ontology.

Table 9
Evaluation results.

Discovered relations	Recall	Precision
علاقة ملكية (has-a)	0.68	0.84
علاقة جنس من (subClassOf)	0.73	0.92

judicial organization; criminal proceedings; etc. Our ontology can also identify these relationships but differently as it is the case in: علاقة ملكية (وكيل الجمهورية, مساعد) also جنس من (محكمة مدنية, محكمة).

However, the missing taxonomic relations in CrimAr ontology relate all the concepts with their appearance in parts of texts. This missing was due to the fact that document structure was not taken into account in the conceptualization as it is the case in ALO ontology.

6.3. Discussion

In this paper, the problem of ontological relation extraction between Arabic compound concepts was addressed by studying their internal structures from which we derived a set of rules. The obtained relations were too general. To specify them, we exploited the semantics of prepositions that allowed us to capture some implicit relations beyond a gamification mechanism.

Facing three different resources (concepts, their properties and relations), we applied FCA/RCA techniques as a powerful formal framework to integrate these heterogeneous resources. The result was a concept lattice translated into ontology coded in DL.

In contrast to other researches based on verbs, we considered that, from the compounds, we can deduce relations linking their components. The deduced relations highly depend on the structure of the compound (N + N, N + Adj, etc.). Thus, we can extract an infinite number of relation instances without being limited to a given type of verbal relation. Some ambiguities can arise when more than one possible relation exists within the same pair of concepts. We overcame this limitation by gamification and through presenting highly specific preposition semantics to players asked to identify the most corresponding one.

Based on the defined rules, our approach achieved encouraging results. Although it has promising performance in terms of precision and recall, our process will be more interesting if we can predict other peculiar compounds whose implicit relation is hard to uncover by a preposition.

Apart from using preposition semantics to deduce implicit relations, we can employ punctuation marks and verbs semantics that make explicit the hidden relations between the compounds. For instance, a comma, when presented between two concepts indicate the presence of a relation.

The empirical results indicated that our proposed approach is efficient for expliciting ontological relations among Arabic compounds concepts. The precisions of 84% and 92% were explained by the fact that we found some incorrect detected relations. The recall, whose values are 68% and 73%, showed that we had undetected true relations. The incorrect matches were due to error rule application, which generally occurs when the identified pattern of the compound nouns is incorrect. This case was caused by syntactic phenomena of the Arabic language. On the other hand, the comparison to a manually-built ontology demonstrated that practically the same concepts existed. They mainly differ in how ontological relations were built and how implicit information was recovered. CrimAr ontology relates the majority of concepts by hierarchical relations to indicate the supervision authorities according to the competencies for the natural person, judicial organization, criminal proceedings, etc. Indeed, ALO ontology can also identify these relationships differently.

The cleanness of the ontology is not so important for the ontology construction task. Although these relations were not included in the final ontology, they are often useful to give an insight about the domain itself and to guide the ontology construction process. Therefore, we should concentrate on increasing the recall of the relation extraction process even at the expense of its precision. The obtained results revealed better values if the patterns are constructed more precisely.

In order to compare the efficiency of our approach with that of the existing works, we must assess the performance of all the approaches when applied on the same corpus for the same relations. Nevertheless, the relations extracted by Sadek and Meziane (2016) are causal, those extracted by Al-Yahya et al. (2016) are antonym and relations extracted by Falih and Omar (2015) are grammatical. Boujelben et al. (2014a) and Boujelben et al. (2014b)) predicted the positions of words which express a semantic relation within a clause, especially the verbal relations. The only similar work study reported in the state-of-the-art is that of Al Zamil and Al-Radaideh (2014).

The highest overall performance averages of their approach on the Newspapers dataset in terms of precision and recall were 89.77% and 84.49%, respectively. Our empirical results outperformed those obtained by Al Zamil and Al-Radaideh (2014) with 2.23% in terms of precision to reach 92%. They slightly decreased in terms of recall due to datasets dissimilarity.

An important implication of our study derives from our result of the learned ontology that will be employed in text retrieval system. We seek to offer the users an opportunity to query legal documents based on a powerful legal ontology, well suited for concept-based information retrieval, to obtain precise results likely to meet their needs.

7. Conclusion

In this paper, we presented our approach to extract relations from Arabic legal text.

We relied on the fact that the elements of a compound are semantically related to each other and if this relation becomes visible, we can decipher a lot of information that can be used as the basis for ontology. For this reason, our approach was initiated by defining a set of rules patterns from compounds concepts. These rules allowed the inducing of general relations. In order to specify them, a gamififcation mechanism was then used based on prepositions semantics. Finally, FCA/RCA techniques were applied to model the concepts hierarchical and transversal relations in order to obtain a lattice concept transformed into an ontology coded in DL.

To evaluate our approach, we compared the derived ontology to a human modeled ontology. Obviously, our approach is efficient in terms of precision and recall. For future work, we intend to extract synonymy relations using linguistic patterns. Similarly, we plan to apply our approach in other fields to prove that it can be efficiently used in various domains.

References

Aguado de Cea, G., Gómez-Pérez, A., Montiel-Ponsoda, E., Suárez-Figueroa, M.C., 2009. Using linguistic patterns to enhance ontology development. In: Proceedings of the International Conference on Knowledge Engineering and Ontology Development. KEOD, pp. 206–213.

Al-Yahya, M., Al-Malak, S., LuluhAldhubayi, 2014. A Pattern-based Approach to Semantic Relation Extraction Using a Seed Ontology. In: ICSC. Newport Beach, California, USA, pp. 96–99.

Al-Yahya, M., Al-Malak, S., LuluhAldhubayi, 2016. Ontological lexicon enrichment: the badea system for semi-automated extraction of antonymy relations from Arabic language corpora. Malaysian J. Comput. Sci. 29 (1), 56–73.

Al Zamil, M.G.H., Al-Radaideh, Q., 2014. Automatic extraction of ontological relations from Arabic text. J. King Saud Univ. – Comput. Inf. Sci. 26 (4), 462–472.

Augenstein, I., Maynard, D., Ciravegna, F., 2014. Relation extraction from the web using distant supervision. In: 19th International Conference on Knowledge Engineering and Knowledge Management EKAW, pp. 26–41.

Bendaoud, R., Hacene, A.M.R., Napoli, A., Toussaint, Y., Delecroix, B., 2007. Text-based ontology construction using relational concept analysis. In: International Workshop on Ontology Dynamics IWOD, pp. 154–163.

Bendaoud, R., Hacene, A.M.R., Napoli, A., Toussaint, Y., Valtchev, P., 2008. Ontology learning from text using relational concept analysis. In: International MCETECH Conference on e-Technologies MCETECH, pp. 154–163.

Bouhriz, N., Benabbou, F., Benlahmer, H., 2015. Text conceptsextraction based on Arabic wordnet and formal concept analysis. Int. J. Comput. App. 111 (16), 30–34.

Boujelben, I., Jamoussi, S., Hamadou, A.B., 2014a. A hybrid method for extracting relations between Arabic named entities. J. King Saud Univ. – Comput. Inf. Sci. 26 (4), 425–440.

Boujelben, I., Jamoussi, S., Hamadou, A.B., 2014b. Relane: discovering relations between Arabic named entities. In: Text, Speech and Dialogue - 17th International Conference. Czech Republic, TSD. Brno, pp. 233–239.

de Bessé, B., Nkwenti-Azeh, B., Sager, J.C., 1997. Glossary of Terms Used in Terminology. vol. 4.

Deterding, S., Sicart, M., Nacke, L., O'Hara, K., Dixon, D., 2011. Gamification. using game-design elements in non-gaming contexts. In: Extended Abstracts on Human Factors in Computing Systems CHI, pp. 2425–2428.

Devisree, V., Reghu Raj, P.C., 2016. A hybrid approach to relationship extraction from stories. Procedia Technol. 24, 1499–1506.

Do, H.H., Melnik, S., Rahm, E., 2003. Comparison of schema matching evaluations. In: NODe 2002 Web and Database-Related Workshops on Web, Web-Services, and Database Systems, London, UK, pp. 221–237.

Dolques, X., Le Ber, F., Huchard, M., Nebut, C., 2013. Analyse relationnelle de concepts pour l'exploration de données relationnelles. In: Conférence Francophone sur l'Extraction et la Gestion des Connaissances EGC, pp. 121–132.

Falih, M.A., Omar, N., 2015. A comparative study on Arabic grammatical relation extraction based on machine learning classification. Middle-East J. Sci. Res. 23 (6), 1222–1227.

Ganter, B., Stumme, G., Wille, R., 2005. Formal Concept Analysis: Foundations and Applications (Lecture Notes in Computer Science / Lecture Notes in Artificial Intelligence). Springer-Verlag New York Inc, Secaucus, NJ, USA.

Ganter, B., Wille, R., 1997. Formal Concept Analysis: Mathematical Foundations. Springer-Verlag New York Inc, Secaucus, NJ, USA.

Gruber, T.R., 1995. Toward principles for the design of ontologies used for knowledge sharing. Int. J. Hum. Comput. Stud. 43 (5–6), 907–928.

Hearst, M., 1992. Automatic acquisition of hyponyms from large text corpora. In: 14th conference on Computational linguistics. Association for Computational Linguistics, pp. 539–545.

Huchard, M., Hacene, A.M.R., Valtchev, P., Roume, C., 2007. Relational concept discovery in structured datasets. Ann. Math. Artif. Intell. 49 (1–4), 39–76.

Huchard, M., Napoli, A., Hacene, A.M.R., Valtchev, P., 2003. Mining description logics concepts with relational concept analysis. In: Brito, P., Bertrand, P., Cucumel, G., Carvalho, F.D. (Eds.), Selected Contributions in Data Analysis and Classification, Studies in Classification, Data Analysis, and Knowledge Organization. Berlin, pp. 259–270.

Huchard, M., Napoli, A., Hacene, A.M.R., Valtchev, P., 2011. A gentle introduction to relational concept analysis, tutorial icfca. In: 9th International Conference on Formal Concept Analysis ICFCA. Nicosia, Cyprus.

Joseph, S., Jasminea, Smitha M., Sheenaa, Na,, 2016. Automatic extraction of hypernym & meronym relations in english sentences using dependency parser. In: 6th International Conference On Advances In Computing & Communications, ICACC. Cochin, India, pp. 539–546.

Kawtrakul, A., Suktarachan, M., Imsombut, A., 2004. Automatic thai ontology construction and maintenance system. In: OntoLex Workshop on LREC.

Kumova, B., 2015. Generating ontologies from relational data with fuzzy-syllogistic reasoning. In: Beyond Databases Architectures and Structures (BDAS). Communications in Computer and Information Science (CCIS), pp. 21–32.

Lakhfif, A., Laskri, M.T., 2016. A frame-based approach for capturing semantics from Arabic text for text-to-sign language mt. Int. J. Speech Technol. 19 (2), 203–228.

Litkowski, K.C., 2002. Digraph analysis of dictionary preposition definitions. In: Workshop on Word Sense Disambiguation: Recent Successes and Future Directions WSD, pp. 9–16.

Maedche, A., Staab, S., 2004. Ontology Learning, HandBook on Ontologies. Springer, International Handbooks on Information Systems.

Mezghanni, I.B., Gargouri, F., 2014. Learning of legal ontology supporting the user queries satisfaction. In: 13th IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT). Warsaw, Poland, pp. 414–418.

Mezghanni, I.B., Gargouri, F., 2015. Towards an Arabic legal ontology based on documents properties extraction. In: 12th IEEE/ACS International Conference of Computer Systems and Applications. AICCSA, Marrakech, Morocco, pp. 1–8.

Mezghanni, I.B., Gargouri, F., 2016. Detecting hidden structures from Arabic electronic documents: Application to the legal field. In: 14th IEEE International Conference on Software Engineering Research, Management and Applications. SERA, pp. 75–81.

Miao, Q., Zhang, S., Zhang, B., Meng, Y., Yu, H., 2012. Extracting and visualizing semantic relationships from chinese biomedical text. In: Pacific Asia Conference on Language, Information and Computation, pp. 99–107.

Pustejovsky, J., 1991. The generative lexicon. Comput. Linguistics 17 (4), 409–441.

Sadek, J., Meziane, F., 2016. Extracting Arabic causal relations using linguistic patterns. ACM Trans. Asian Low-Resour. Lang. Inf. Process. 15 (3), 14:1–14:20.

Sag, I., Baldwin, T., Bond, F., Copestake, A., Flickinger, D., 2002. Multiword expressions: a pain in the neck for NLP. In: Proceedings of the Third International Conference on Intelligent Text Processing and Computational Linguistics CICLing, London, UK, pp. 1–15.

Sruti Rallapalli, S.P., 2012. A hybrid approach for the interpretation of nominal compounds using ontology. In: 26th Pacific Asia Conference on Language, Information and Computation PACLIC, pp. 554–563.

Ta, C.D., Thi, T.P., 2016. Automatic Extraction of Semantic Relations from Text Documents. Can Tho City, Vietnam, pp. 344–351.

Takase, S., Okazaki, N., Inui, K., 2015. Fast and large-scale unsupervised relation extraction. In: Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation, PACLIC. Shanghai, China, pp. 96–105.

Vela, M., Declerck, T., 2009. A methodology for ontology learning: deriving ontology schema components from unstructured text. In: Workshop on Semantic Authoring, Annotation and Knowledge Markup.

Xiang, Y., Chen, Q., Wang, X., Qin, Y., 2016. Distant supervision for relation extraction with ranking-based methods. Entropy 18 (6), 204.