



A Bottom-up Approach for Moroccan Legal Ontology Learning from Arabic texts

MISC laboratory, Faculty of Science Ibn Tofail University
Kenitra, Morocco

Kaoutar BELHOUCINE

Aziz MOULOUDI

Mohammed MOURCHID

Samir MBARKI

Summary



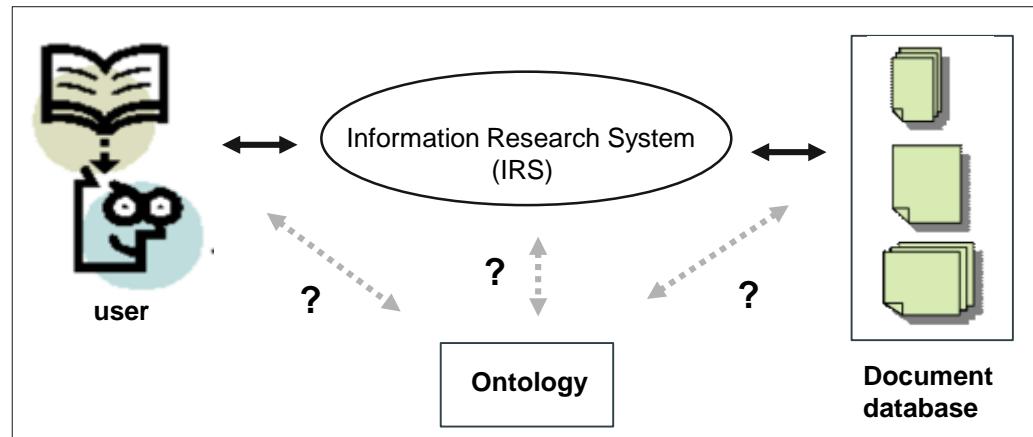
1. Introduction
2. Ontology Learning from Texts
3. Our approach, results
4. Conclusion and perspective

Introduction

Motivation

One of the problems of the current indexing of documents by words, or groups of words, is not to take into account the language phenomena which are synonymy and homonymy:

- **Synonymy** generates documentary silence,
- **Homonymy** generates documentary noise.



→ The solution consists in indexing the documents by concepts, while preserving the links: concept / term (s).

Ontology Learning from Texts

Ontology Learning Layer Cake

$\forall x,y,z \text{ (applied_to}(x,y), \text{similar_to}(y,z))$
 $\rightarrow \text{applied_to}(x,z))$

Represents (dom:Lawyer, range:client)

Is_a(LAWYER, PERSON)

LAWYER, PERSON

{infraction, crime}

case, infraction, crime

Axioms

Relation

Concepts hierarchies

Concepts

Synonyms

Terms

Ontology Learning Tools

Tools	Language	Elements extracted	Techniques
Terminae (2005)	French	Terms, synonyms, concepts, taxonomies, non-taxonomic relations	Linguistic and knowledge engineering
Text2Onto (2005)	English Spanish German	Terms, synonyms concepts, taxonomies, non-taxonomic relations, instances	linguistic processing statistical text analysis machine learning association rules
OntoGain (2006)	English	Terms, concepts, taxonomies	Machine learning, text mining

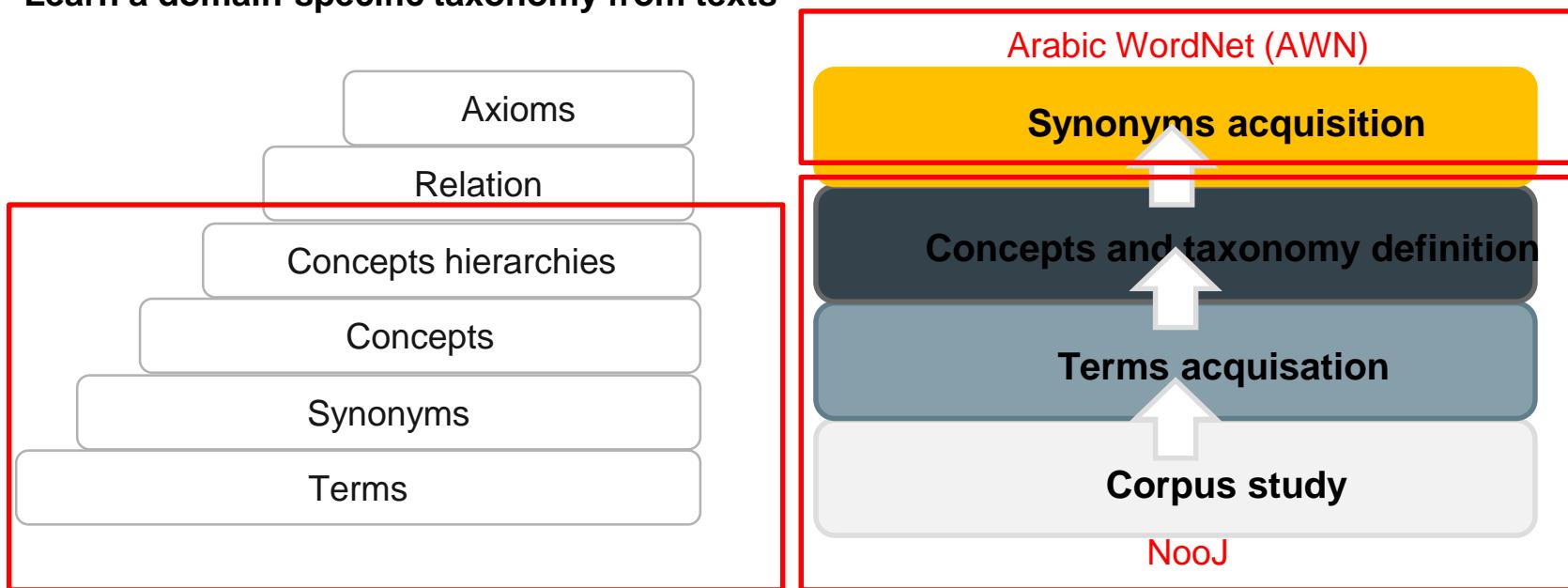


The Arabic language remains not supported by these tools, even it is one of the most spoken languages in the world.

Our Approach

Bottom-up approach

Learn a domain-specific taxonomy from texts



Corpus Study

The Moroccan Family Code : 7 books, 400 Articles in Arabic langage.

File Name	Size	Last Modif.
kitab1	294104	14/05/2020
kitab2	274376	14/05/2020
kitab3	184142	14/05/2020
kitab4	382983	14/05/2020
kitab5	78623	14/05/2020
kitab6	268654	14/05/2020
kitab7	22872	14/05/2020

Corpus Language is "Arabic(ar)";
Original Text File format is "WORD".
Corpus consists of 7 text files
Text Delimiter is: "\n"
Corpus contains 2681 text units delimited by "\n"
1607 digits
1035 other delimiters

Characters
Tokens
Digrams
Ambiguities
Unambiguous Words

Remove Add

الكتاب الأول: الزواج

الفصل الأول: الخطبة والزواج

المادة 4

الزواج ميثاق تراضي وترابط شرعي بين رجل وامرأة على وجه الدوام، غالباً الإحسان والعفاف وإنشاء أسرة مستقرة، برعاية الزوجين طبقاً لأحكام هذه المدونة.

الباب الأول: الخطبة

المادة 5

الخطبة تواعد رجل وامرأة على الزواج.

تحقق الخطبة بتعبير طرقها بأي وسيلة متعارف عليها تفيد التواعد على الزواج، ويدخل في حكمها قراءة الفاتحة وما جرت به العادة والعرف من تبادل الهدايا.

المادة 6

يعتبر الطرفان في فترة خطبة إلى حين الإشهاد على عقد الزواج، وكل من الطرفين حق العدول عنها.

Dictionnary Definition

Based on the Bilingual Legal Lexicon of the Family Code (French Arab).

<http://adala.justice.gov.ma/production/OutilsPratiques/fr/LexiqueFamilleFR/index.htm>

Category	Code
Noun	N
Adjective	ADJ
Adverb	ADV
Personal pronoun	PRON
Demonstrative	DEM
Relative	REL
Preposition	PREP
conjunction	CONJ

800 entries

Lexique juridique bilingue du Code de la famille
français-arabe
Par
Meryem Serraj – Karima Arbia – Khadouja Ferdani

Le lexique juridique bilingue du Code de la famille français-arabe est une base de données établie à partir du vocabulaire officiel de la loi n° 70-03 portant Code de la famille promulguée par le Dahir n° 1-04-22 du 3 février 2004 (12 Hjâja 1424).

Ce lexique juridique permet de consulter :

- la traduction en langue arabe de 703 concepts, termes et locutions contenus dans le Code de la famille pour illustrer le sens de chaque mot ;
- la définition lexicographique et terminologique des notions doctrinaires relatives au droit musulman ;
- les articles dans lesquels le mot, le concept ou l'expression figure.

Cliquez dans l'alphabet puis sur le mot de votre choix

A B C D E F G H I J K L M N O P R S T U V W
A

Abandon du domicile conjugal
Abandon de famille

```
# NooJ V5
# Dictionary
3
4 # Language is: ar
5 # Alphabetical order is not required.
6 # Use inflectional & derivational paradigms' description file
7 # Special Command: #use paradigms.nof
8
9 # Special Features: +NW (non-word) +FC (frozen expression code)
10 # +FLX= (inflectional paradigm) +DRV= (derivation paradigm)
11 # +S= (synonym) +D= (definition) +P= (part-of-speech)
12 # +C= (category) +L= (lexeme) +M= (morpheme)
13 # +O= (orthographic form) +R= (rule)
14 # Special Characters: '\ " ' ' ' ' ' ' ' ' ' ' '
15 #
16
17 #use famille_v4.nof
18
19 ##### nom primitif #####
20 #+N+Jur+FLX=FLxAB+DRV=Abae:Flx
21 #+N+Jur+FLX=FLxM+DRV=Abnae:Flx
22 #+N+Jur+FLX=FLxM+DRV=Abnaah:Flx
23 #+N+Jur+FLX=FLxM+DRV=Ikhbaa:Flx
24 #+N+Jur+FLX=FLxM+DRV=Ikhbaah:Flx
25 #+N+Jur+FLX=FLxM+DRV=DrvM
26 #+N+Jur+FLX=FLxF+DRV=Osara:Flx
27 #+N+Jur+FLX=FLxF+DRV=Abnae:Flx
28 #+N+Jur+FLX=FLxF+DRV=Abnaah:Flx
29 #+N+Jur+FLX=FLxF+DRV=Onahati:Flx
30 #+N+Jur+FLX=FLxF+DRV=Omam:Flx
31 #+N+Jur+FLX=FLxF
32 #+N+Jur+N
```

Dictionary Definition

Inflected and derived forms

Code	Meaning
a	Accusative
u	Nominative
i	Genitive
an	Tanwin, Accusative
un	Tanwin, Nominative
ins	Tanwin, Genitive
m,f	Male, Female
S,d, p	Singular, dual, plural

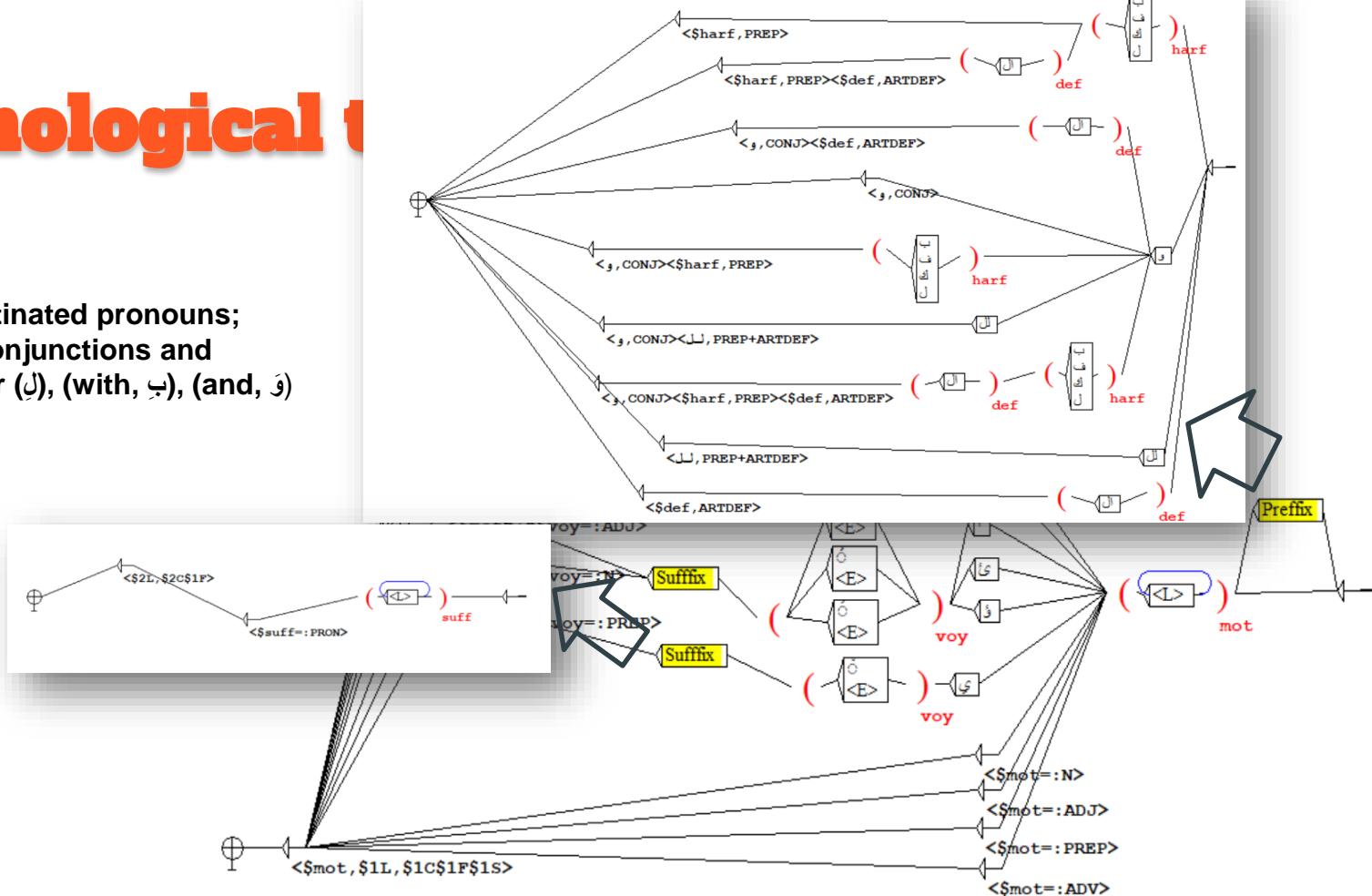
```

1 # Nooj V5
2 # Inflectional/Derivational description
3 #
4 # Language is: ar
5 #
6 # Generic Commands:
7 # <B>: keyboard Backspace
8 # <C>: change Case
9 # <D>: Duplicate current char
10 # <E>: Empty string
11 # <L>: keyboard Left arrow
12 # <N>: go to end of Next word form
13 # <P>: go to end of Previous word form
14 # <R>: keyboard Right arrow
15 # <S>: delete/Suppress current char
16 # Arguments for commands <B>, <L>, <N>, <P>, <R>, <S>:
17 # xx number: repeat xx times
18 # W: whole word
19 # Examples
20 # <R3>: go right 3 times
21 # <LW>: go to beg. of word
22 #
23 # Language-Specific Commands:
24 # <T>: replaces Teh marbuta with Teh
25 # <M>: processes consonant 'n' for past form, 3rd person, feminine plural
26 # <Z>: processes consonant 't' for past form, 1st person, singular
27 #
28 # Special Characters: '=' '<' '>' '\'' '\"' ':' ';' '|'+ '-' '/' '$' '_' ',' '#'
29 #
30 #
31 DrvM=
32 #Les formes au masculin pluriel
33 نـونـ /N+m+pt+a /
34 نـونـ /N+m+pt+u /
35 نـونـ /N+m+pt+i /

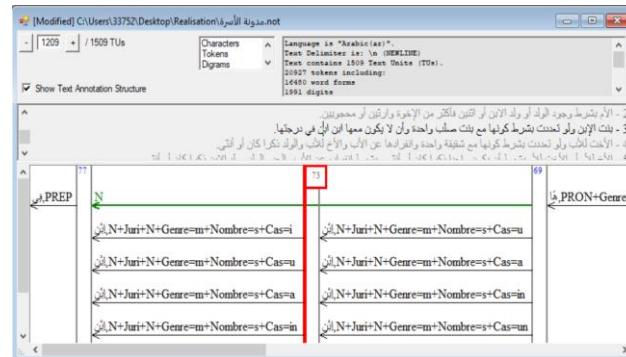
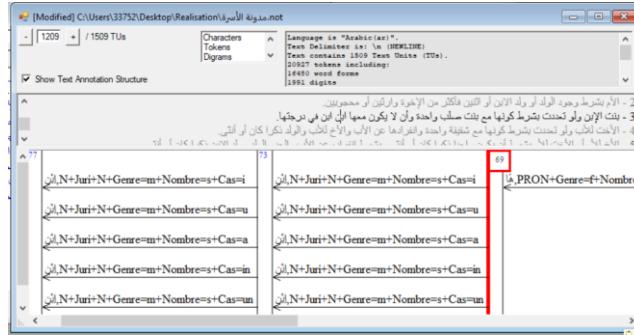
```

Morphological t

- Definite article;
 - Personal agglutinated pronouns;
 - Agglutinated conjunctions and prepositions for (لـ), (with, بـ), (and, وـ)



Ambiguity treatment



Son

ابن

daughter

بنت

cousin

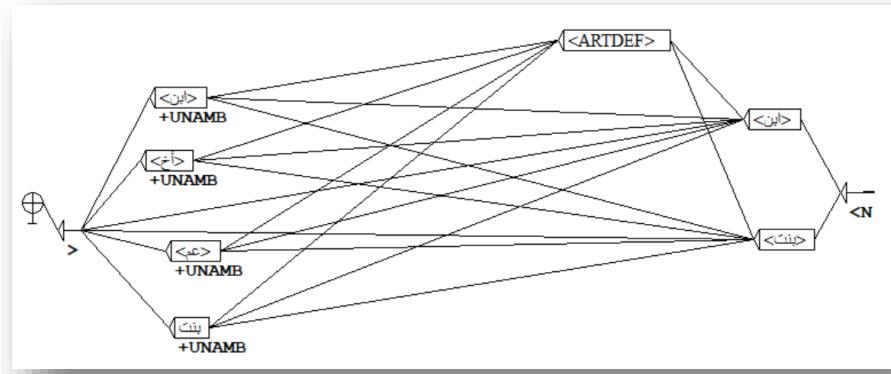
ابن العم

Son of son

ابن ابن

nephews

ابناء الإخوة



Dictionary Definition

Semantic grouping

+MARRIAGE

زواج, Marriage
إيجاب, offer
قبول, acceptance
صدقان, dowry
عُقد زواج, marriage contract
جهاز, the furniture

+ DISSOLUTION

فسخ, Annullment
طلاق, Divorce
تطليق, Repudiation
غياب, Absence
شقاق, Irreconcilable Differences

+INHERITANCE

حجب
تفضيل
إرث, inheritance
تركة, estate
فرض
أكدرية

+BIRTH

بنوة, filiation
نسب, Paternity
فراش, the conjugal bed
حضانة, the custody
أمومة, maternity
ولادة, birth

+ WILLS

موصي,
وصية,
تنزيل,
الموصى به,

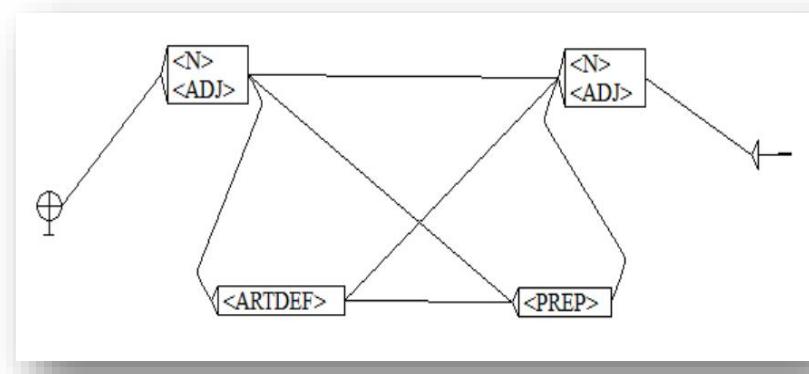
+CAPACITY

أهلية, legal capacity
حُجْر, Guardianship
ولي, legal tutor
وصي, testamentary guardian
نيابة الشرعية, Legal Representation

Terms acquisition

We manually extract a list of about 13 patterns of nominal compositions.

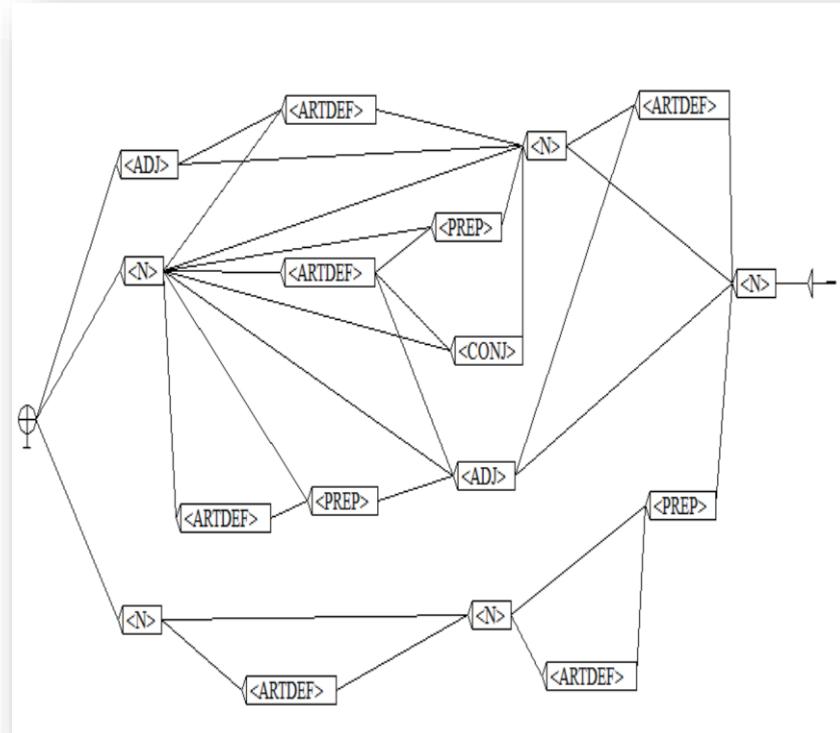
Pattern	Example
N	الزواج, Marriage الخطبة, Engagement
N1_N2	عقد الزواج, Marriage contract عدة الطلاق, Legal waiting period (iddat) following a revocable divorce
N1_ADJ	حجز تحفظي غرامة تهديدية
ADJ_N1	عديم الأهلية ، Person devoid of capacity ناقص الأهلية ، Person Not Fully Capable
N1_prepN2	التواعد على الزواج الإشهاد على عقد
N1_prepADJ	زواج غير صحيح عدة غير الحامل نسب غير شرعي



Terms acquisition

We manually extract a list of about 13 patterns of nominal compositions.

Pattern	Example
N1_N2_N3	فسخ عقد الزواج Annulment of the marriage contract وثيقة عقد الزواج
N1_N2_ADJ	بنوة ولد مجهول The filiation of a child of unknown paternity شخص غيروارث
N1_ADJ_N2	شخص عديم الأهلية Person devoid of capacity طفل مجهول النسب
N1_N2_conj_N3	لأحكام الولاية أو الوصاية Provisions governing legal tutelage, testamentary guardianship or courtappointed guardianship
N1_N2_prepN3	طلاق الحلف بالحرام
N1_ADJ_prepN2	المصالح المالية لقاصر نيابة شرعية عن قاصر
N1_perpN2_N3	إشهاد على عقد الزواج



Terms acquisition

We manually extract a list of about 13 patterns of nominal compositions.

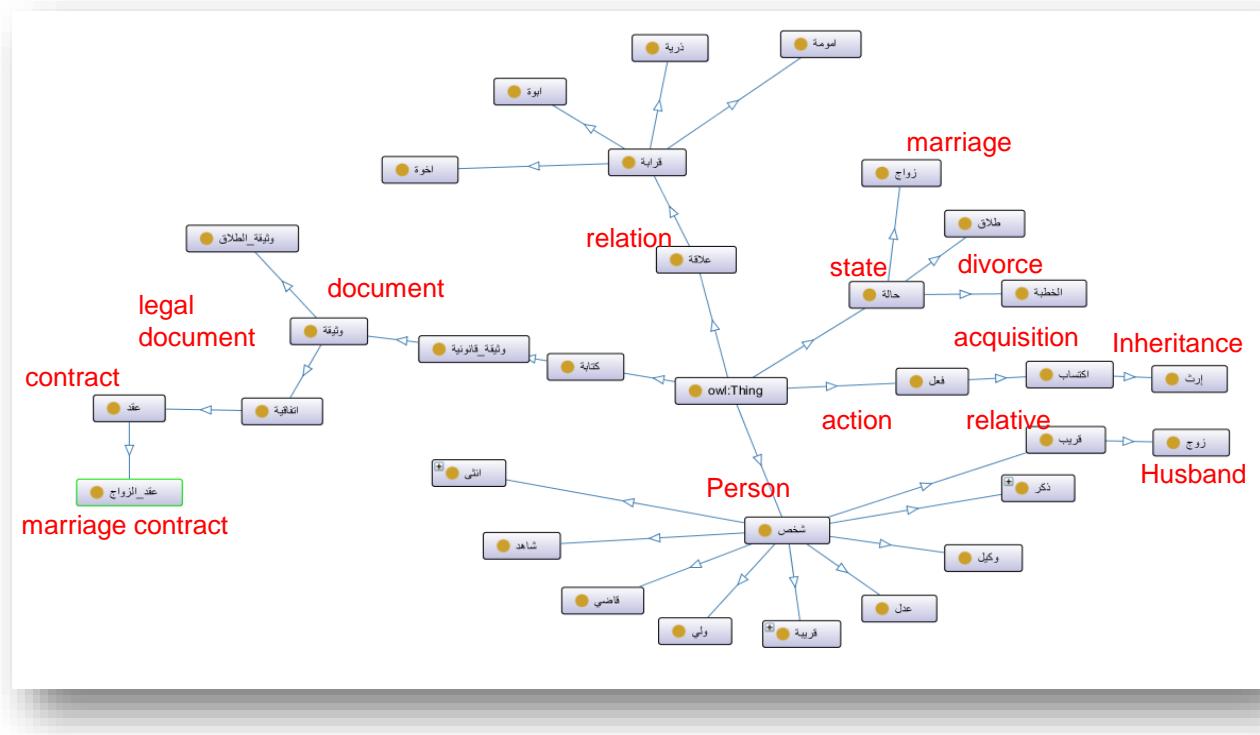
The screenshot shows a software interface for linguistic analysis, specifically for extracting patterns of nominal composition. The main window displays a grid of word frequencies across seven different corpora (kitab1 to kitab7). The grid highlights specific patterns in yellow, such as 'الزوج النكرة' (the male noun) and 'النكرة والزوج الماءدة 4 الزوج مثلاً في ترتيب وترتبه' (the noun and its antecedent in the fourth position, such as in the sequence 'zawj al-nukra'). The bottom left of the interface shows a 'Query' window with the text 'الزوج النكرة' and the count '1014/1014'. The bottom right lists some of the extracted patterns: 'ادارة المؤمن', 'ادن المحكمة', 'ازلة الاصح', 'ازلة اصحاب', and 'اسرة اصحاب'.

Concepts and taxonomy definition

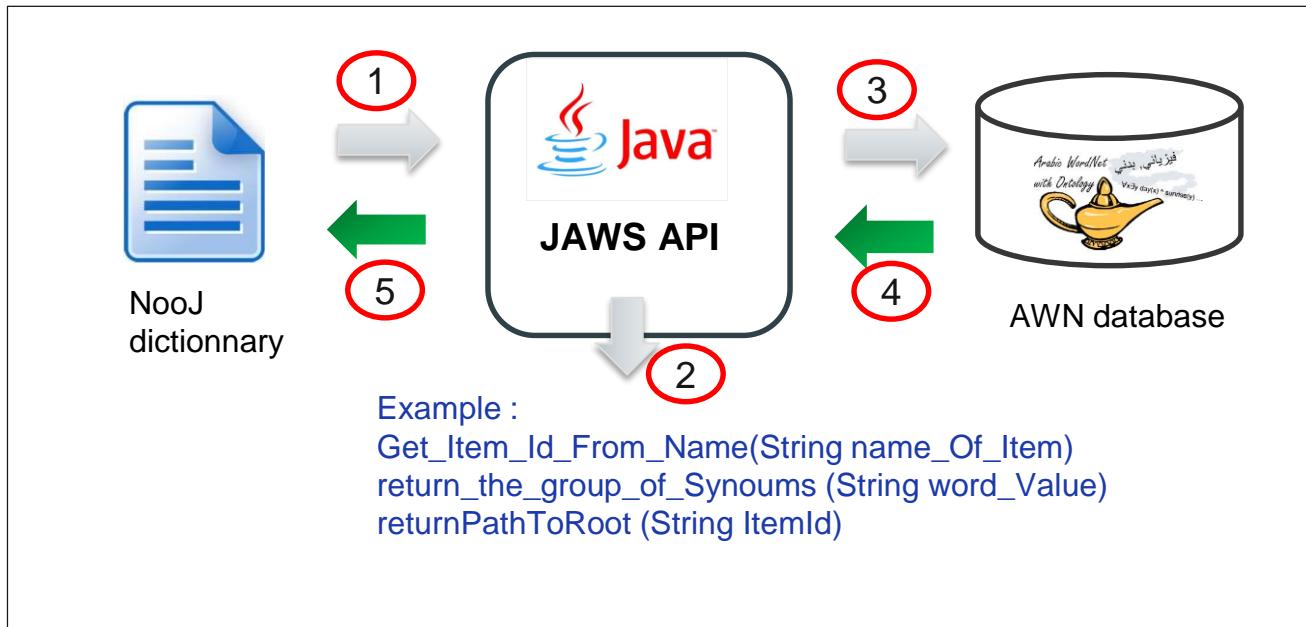
<p>إقرار حق الطلاق الفسخ أو الطلاق التعبير عن الطلاق بإشهاد على الطلاق أسباب الطلاق بتوثيق الطلاق</p>	<p>To register the marriage الزواج رسم</p>	<p>مستندات الزواج أهلية الزواج بيطان الزواج بتوثيق الزواج</p>	<p>بتوثيق عقد الزواج رد عقد الزواج فسخ عقد الزواج ملف عقد الزواج وثيقة عقد الزواج وكالة عقد الزواج انحلال عقد الزواج ابرام عقد الزواج فسخ عقد الزواج شروط عقد الزواج الإرادية لعقد الزواج</p>	<p>نفقة القرابة kinship maintenance نفقة المحسضون child's maintenance نفقة نفسه نفقة الزوجة Maintenance Paid to the Wife نفقة الأقارب Maintenance Paid to Close Relatives نفقة الأب نفقة الابنة نفقة الآباء</p>	<p>أحكام الإسلام السمحاء أحكام المذهب المالكي أحكام قانون الأحوال أحكام هذه المدونة أحكام العقد ومقاصده أحكام الحضانة أحكام الوصية أحكام عامة أحكام إهمال الأسرة أحكام مسؤولية الوكيل أحكام انتقالية وختامية أحكام الطهير الشريف أحكام الكتاب أحكام القسم أحكام الباب أحكام المادة أحكام الفصل</p>
<p>To register the divorce وثيقة الطلاق حق إيقاع الطلاق أنواع الطلاق</p>	<p>The marriage decree عقد الزواج ولي الزواج لسن الزواج فترقة الزواج الولاية في الزواج أنواع الزواج</p>	<p>Marriage Categories إثبات الزواج تاریخ الزواج</p>			
<p>Divorce Categories بانقضاء عدة الطلاق تاريخ الطلاق عدة الطلاق رسم الطلاق</p>					
<p>The divorce decree</p>					

- Divorce and marriage share several syntactic context
 - Close relative, child's, wife, kinship share the same syntactic context

Concepts and taxonomy definition



Synonyms acquisition



Synonyms acquisition

Concept	Reference Tree	Synset	Synset tree	Similarity
زوج Husband	زوج -> قريب-> شخص Husband →relative->person-	شريك حياة، خليل، بغل، قرين، زوج Spouse, partner, married person, mate, better half	زوج -> قريب-> شخص-> مسيب -> كيئونة Spouse->relative->person->organism->living_thing->object_entity	3
		زوج Pair	زوج->جتمع->جماعة->جماعة Pair -> gathering->social_group->group	1
		زوج Pair, brace	زوج->مجموعة->مجموعة->جماعة Pair->set->collection->group	1
		زوج، زوجة Wife, married woman	زوجة->إمرأة->أنثى->شخص-> مسيب -> كيئونة Wife->woman->female->person->organism->living_thing->object_entity	0
		عاشق، حبيب، محظوظ، قرين، رجل، خليل، زوج man	رجل->ذكر->شخص-> مسيب -> كيئونة Man->male->person->organism->living_thing->object_entity	0
إرث Inheritance	إرث->اكتساب-> فعل Inheritance->acquisition->action	إرث، ميراث، تركة، وراثة، ورثة Inheritance, heritage	إرث->اكتساب->إحراز-> عمل->إنجاز-> عمل-> فعل Inheritance->acquisition->acquiring->deed->accomplishment->action->act	3

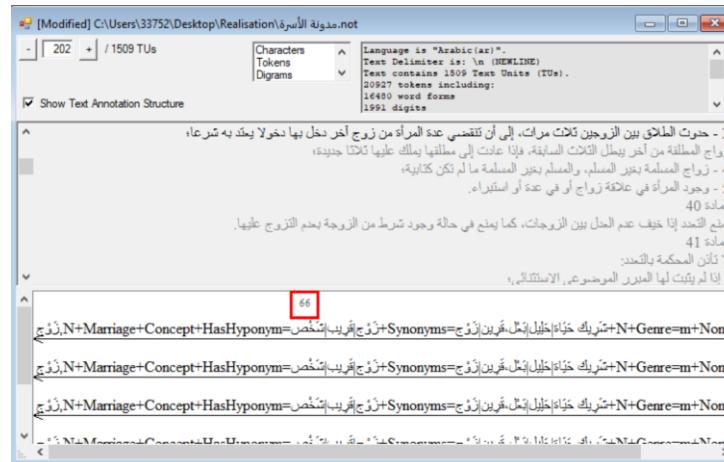
Results

Concepts and taxonomy definition

Concept, GrammaticalCategory+Concept+SemanticGrouping+HypernymTree=list_of_String+Synonyms=list_of_String

Example

زوج|فريـن|بـعـل|خـليل|شـريكـ حـيـاة=N+Concept+Marriage+HypernymTree=Synonyms
زوج|فـريـبـ|شـخص=N+Concept+Inheritance+HypernymTree=Synonyms
زـوجـ|أـرـثـ|إـرـثـ=N+Concept+Wills+HypernymTree=Synonyms



Conclusion perspective

Questions ?