

# Arabic Ontology

## الأنطولوجيا العربية

**Mustafa Jarrar**

[Birzeit University, Palestine](#)

[mjarrar@birzeit.edu](mailto:mjarrar@birzeit.edu)

[www.jarrar.info](http://www.jarrar.info)





Watch this lecture and download the slides from  
<http://jarrar-courses.blogspot.com/2011/12/arabic-ontology.html>



# The lecture is based on:

Mustafa Jarrar: **Building A Formal Arabic Ontology** (Invited Paper) . In proceedings of the Experts Meeting On Arabic Ontologies And Semantic Networks. Alecso, Arab League. Tunis, July 26-28, 2011. Article <http://www.jarrar.info/publications/J11.pdf.htm>  
Slides: <http://mjarrar.blogspot.com/2011/08/building-formal-arabic-ontology-invited.html>

Mustafa Jarrar: **Towards The Notion Of Gloss, And The Adoption Of Linguistic Resources In Formal Ontology Engineering**. In proceedings of the 15th International World Wide Web Conference (WWW2006). Edinburgh, Scotland. Pages 497-503. ACM Press. ISBN: 1595933239. May 2006. <http://www.jarrar.info/publications/J06.pdf.htm>

Please use both articles if when citing the Arabic Ontology



## Part 1: The Arabic Ontology Design

- Part 2: Gloss in the Arabic Ontology
- Part 3: The Top Levels of the Arabic Ontology
- Part 4: Arabic Ontology Vs WordNet
- Part 5: Building Synsets Automatically

### Lecture Keywords:

Arabic Ontology, Why Arabic Ontology, Ontology, Linguistic Ontology, Lexical Semantics, Semantics, Meaning, Concept, Upper Level Ontology, Lexical Relation, Semantic Relation, Subtype, subsumption, Hyponymy, Meronymy, Inheritance, Part-whole, WordNet, Arabic WordNet, Synonymy, Polysemy, Gloss, Ontology versus WordNet, Ontology Matching, Thesaurus construction,

الأنطولوجيا العربية، استخدام الأنطولوجيا العربية، الأنطولوجيا اللغوية، علم الدلالة، الدلالة، المعنى، المفهوم، حدود الأنطولوجيا، العلاقات اللغوية، العلاقات المفاهيمية، علاقة الجنس، الوراثة، جزء-كل، شبكة المفردات، شبكة المفردات العربية، تعدد المعاني، المترادف، تعريف الحد، الفرق بين الأنطولوجيا وشبكة المفردات، ربط الأنطولوجيات المفاهيمية، مكنز، بناء المكنز آلياً.

# Application ontology vs. Linguistic Ontology

- The importance of linguistic ontologies is growing rapidly.
  - **An application ontology** is a representation of the semantics of a certain domain/application. Such as, the FOAF ontology, the Palestinian e-government ontology, the CContology, etc.
    - Each word convey one concept (no polysemy).
    - Represents application's knowledge and data structure.
    - Used only by a certain application, or a class of applications.
  - **A linguistic ontology** is a representation the semantics of all words of a human language, independently of a particular application. Such as WordNet for English.
    - Each word may convey several concepts (Polysemy).
    - Represents common-sense knowledge (lexical semantics).
    - Can be used for general purposes.
- ➔ Let's first understand the relations between a word and its meaning(s).

# The Arabic Ontology Project

<http://sina.birzeit.edu/ArabicOntology/>

- A project started in 2010, at Sina Institute, Birzeit University, Palestine.
- The Arabic Ontology is **can be use an Arabic WordNet**, but it is more.
- Unlike WordNet, the ArabicOntology is **logically and philosophically well-founded**, as it follows strict ontological principles. → but can be used an Arabic WordNet.
- Built semi-manually

→ The project is partially funded (Seed funding) by Birzeit University (VP academic Office, Research Committee).

The screenshot shows the website for the Arabic Ontology Project. The main heading is "أنتولوجيا اللغة العربية" (Arabic Ontology) with the English translation "ARABIC ONTOLOGY" below it. A navigation menu includes "About", "People", "Publications", "News & Events", "Downloads", and "Online Search".

The "About Arabic Ontology" section contains the following text: "The Arabic Ontology is a formal representation of the concepts that the Arabic terms convey. For each term in the Arabic language, the set of its meanings (i.e. concepts) are identified, and semantic relationships (such as subtype-of and part-of) between all concepts are introduced. For simplicity, the Arabic ontology is a tree of the meanings of the Arabic terms (see the diagram below). The Arabic Ontology can be seen as an Arabic WordNet; however its relationships are well-formalized, and glosses follow strict formulation and ontological rules."

The "News & Events" section lists the following items:

- The Arabic Ontology Project receives seed funding from Birzeit University
- First Arabic Ontology Volunteers Meeting
- Dr. Jarrar talk about the Arabic Ontology at Birzeit IT Expo 2010

The diagram below the text illustrates the structure of the ontology. It shows a central node "جدول" (Table) with several branches: "مصفوفة" (Matrix), "تصنيف" (Classification), "أجندة" (Calendar), "نهر" (River), "صف" (Page), and "الجدول الدوري" (Periodic Table). Each branch is further detailed with numbered sub-points and brief descriptions.

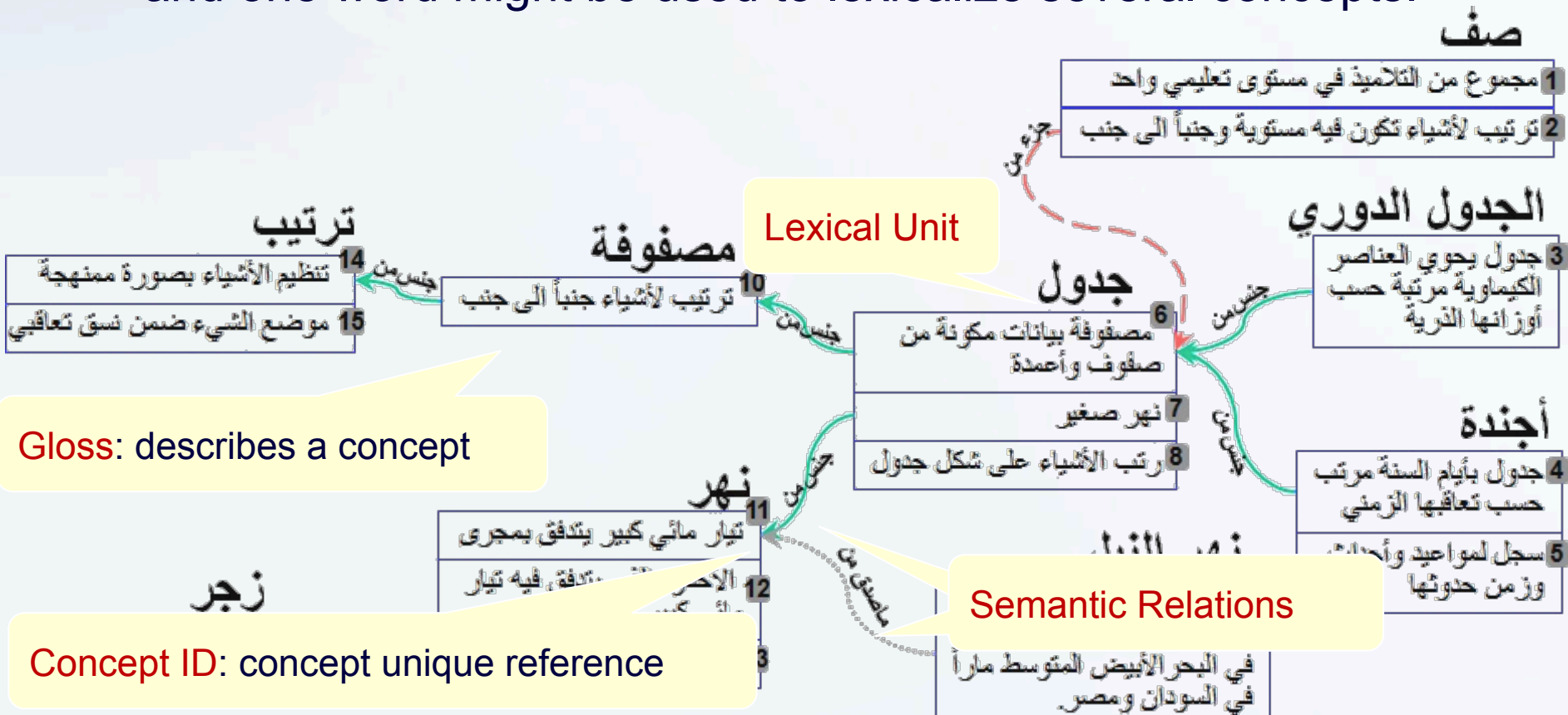
# Why the Arabic Ontology?

In application scenarios such as

- **Information Search and Retrieval** -to enrich queries and improve the quality of the results, i.e. meaningful search rather than string-matching search;
- **Machine Translation and Term Disambiguation** -by finding the exact mapping of concepts across languages, specially that the Arabic ontology is also mapped to the WordNet;
- **Data Integration and Interoperability** -in which the Arabic ontology can be used as a semantic reference to several autonomous information systems;
- **Semantic Web and Web 3.0** -by using the Arabic ontology as a semantic reference to disambiguate the meanings used in the web sites;
- among many, **many other applications.**

# Arabic Ontology: Data Model (Simplified)

- **ConceptID** (as a synsetID in WordNet) to identify a concept.
- **Polysemy and synonymy**: like in WordNet, several words (i.e., lexical units) can be used to lexicalize one concept (synonymy); and one word might be used to lexicalize several concepts.





# Lexical vs. Semantic Relationships

- **Semantic relations** (علاقات مفاهيمية) are relationships between concepts (not words), e.g., subtype, part-of, etc.
- **Lexical relations** (علاقات لغوية) are relationships between words (not concepts), e.g., synonym-of, root-of, abbreviation-of, etc.
- **Ontologies** are mainly concerned with semantic relations.

صف

- 1 مجموع من التلاميذ في مستوى تعليمي واحد
- 2 ترتيب لأشياء تكون فيه مستوية وجنباً إلى جنب

Lexical Unit

الجدول الدوري

- 3 جدول يحوي العناصر الكيماوية مرتبة حسب أوزانها الذرية

جدول

- 6 مصفوفة بيانات مكونة من صفوف وأعمدة
- 7 نهر صغير

أجندة

- 4 جدول بأيام السنة مرتب حسب تعاقبها الزمني
- 5 سجل للمواعيد وأحداث وازمن حدوثها

Semantic Relations

نهر

- 11 تيار مائي كبير يتدفق بمجرى
- 12 نهر

نهر النيل  
نهر في إفريقيا يتبع من بحيرة فيكتوريا وبحيرة تانا ويصب في البحر الأبيض المتوسط ماراً في السودان ومصر.

ترتيب

- 14 تنظيم الأشياء بصورة ممنهجة
- 15 موضع الشيء ضمن نسق تعاقبي

زجر

- 16 عنف بقية الشعب

Gloss: describes a concept

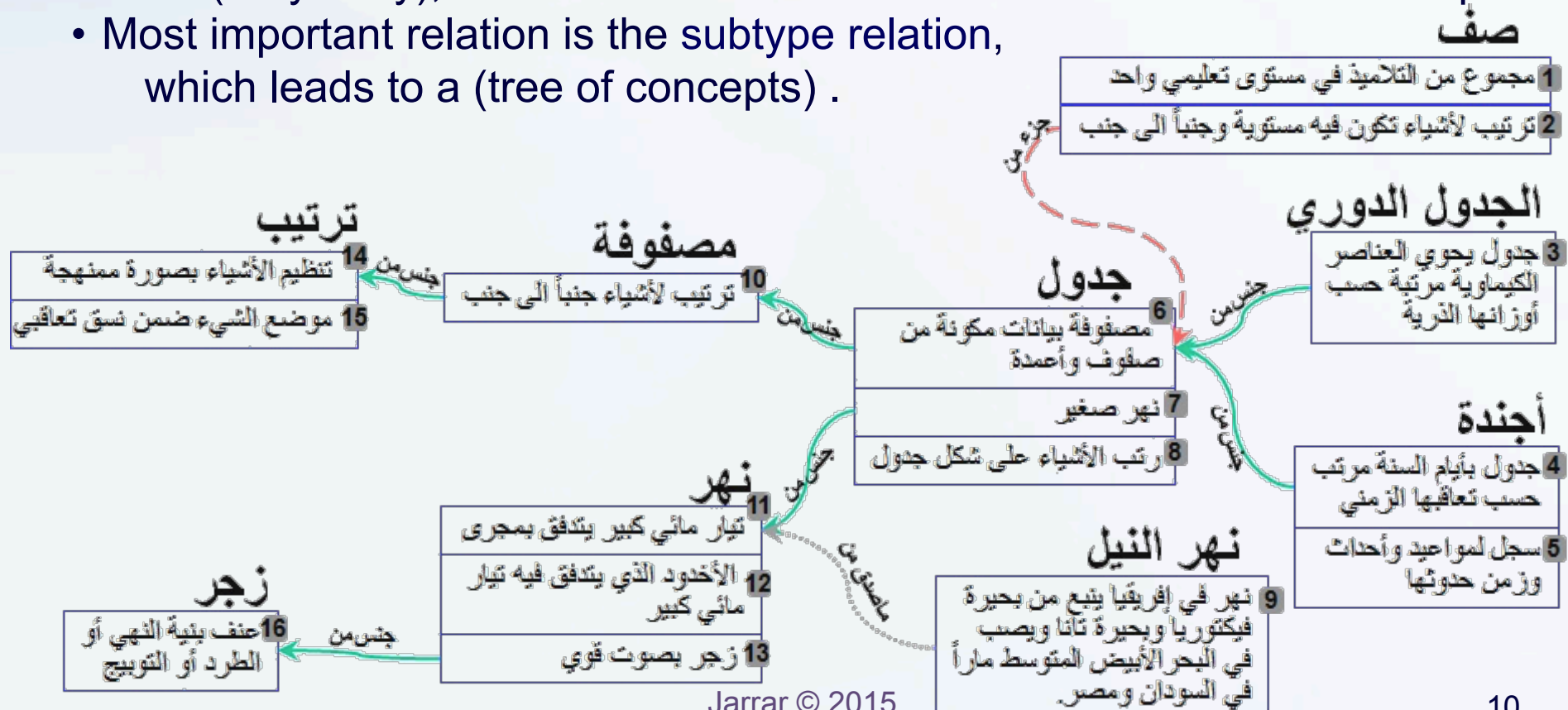
Concept ID: concept unique reference

# Arabic Ontology

- **Arabic Ontology**: the set of concepts (of all Arabic terms), and the semantic (not lexical) relationships between these concepts.

الان طول وجيا ال عربي هي مجموعة معاني لكلمات اللغة العربية، و علاقات بين

- **To build an Arabic Ontology**: Identify the set of concepts for every Arabic word (Polysemy), and define semantic relations between these concepts.
- Most important relation is the subtype relation, which leads to a (tree of concepts) .



# Arabic Ontology: Subtype Relationships

- **Subtype relation** (علاقة جنس من): is a mathematical relations (subset:  $A \subseteq B$ ), such that every instance in A must also be an instance of B.
- **Inheritance** (التوارث): subtypes inherit all properties of their super types.
- “**Hyponymy**” in WordNet is close to (but not the same as) the subtype relation.
- “**General-Specific** (أعم-أخص):” relations, as in thesauri, are not subtype relations.
- **Attention: instance vs Type (Mustafa vs. Person)**

صف

- 1 مجموع من التلاميذ في مستوى تعليمي واحد
- 2 ترتيب لأشياء تكون فيه مستوية وجنباً إلى جنب

جزء من

## الجدول الدوري

- 3 جدول يحوي العناصر الكيماوية مرتبة حسب أوزانها الذرية

## جدول

- 6 مصفوفة بيانات مكونة من صفوف وأعمدة
- 7 نهر صغير
- 8 رتب الأشياء على شكل جدول

جنس من

## أجندة

- 4 جدول بأيام السنة مرتب حسب تعاقبها الزمني
- 5 سجل المواعيد وأحداث وزمن حدوثها

## نهر النيل

- 9 نهر في إفريقيا ينبع من بحيرة فيكتوريا وبحيرة تانا ويصب في البحر الأبيض المتوسط ماراً في السودان ومصر.

جنس من

مماثل من

## مصفوفة

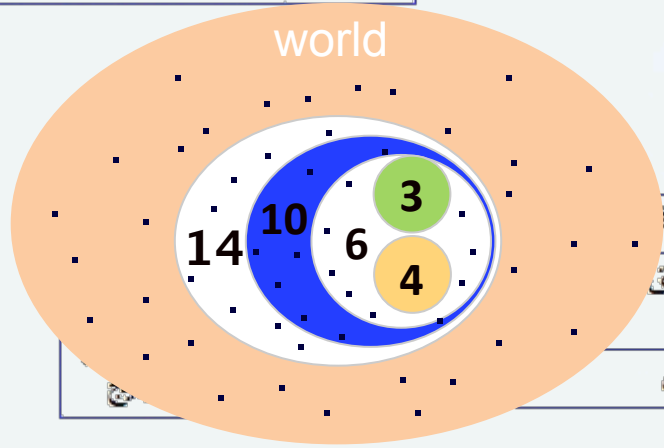
- 10 ترتيب لأشياء جنباً إلى جنب

## ترتيب

- 14 تنظيم الأشياء بصورة ممنهجة
- 15 موضع الشيء ضمن نسق تعاقبي

## نهر

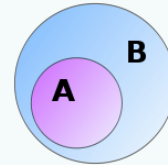
- 11 تيار مائي كبير يتدفق
- 12 الأخدود الذي يتدفق مائي كبير
- 13 زجر بصوت قوي



# Arabic Ontology: Subtype Relationships

- It is recommended to use **proper subtypes**, as it is more strict.
- That is, A and B are never equal, B is always a super set of A.
- It is recommended to classify concepts based on “**rigidity**”.
- For example it is wrong to say that a ‘WorkTable’ is type of ‘Table’.  
as being a work table is a non-rigid property.
- As such, subtypes form a **tree**.

$$A \subsetneq B.$$



وبالتالي تصبح ال اطولوجيا شجرة وليس  
شبكة معاني

صف

- 1 مجموع من التلاميذ في مستوى تعليمي واحد
- 2 ترتيب لأشياء تكون فيه مستوية وجنباً إلى جنب

الجدول الدوري

- 3 جدول يحوي العناصر الكيماوية مرتبة حسب أوزانها الذرية

جدول

مصفوفة بيانات مكونة من صفوف وأعمدة

جنس من

ترتيب

- 14 تنظيم الأشياء بصورة ممنهجة

- 15 موضع الشيء ضمن نسق تعاقبي

مصفوفة

ترتيب لأشياء جنباً إلى جنب

جنس من

أجدد

- 4 جدول حسب

- 5 سجل ل

- وزمن

Please see my lecture about Ontology Modeling (OntoClean)

<http://jarrar-courses.blogspot.com/2012/05/aai-ontocleanavi.html>

- 10 حنف بنية التهي أو الطرد أو التوبيج

جنس من

زجر بصوت قوي

فيحسوريا وبحيره نانا ويصب في البحر الأبيض المتوسط ماراً في السودان ومصر.

- Part 1: The Arabic Ontology Design
-  • Part 2: Gloss in the Arabic Ontology
- Part 3: The Top Levels of the Arabic Ontology
- Part 4: Arabic Ontology Vs WordNet
- Part 5: Building Synsets Automatically

# Glosses in the Arabic Ontology

Writing glosses in the Arabic Ontology follow strict ontological rules (see [J06])

صف

- 1 مجموع من التلاميذ في مستوى تعليمي واحد
- 2 ترتيب لأشياء تكون فيه مستوية وجنباً الى جنب

الجدول الدوري

- 3 جدول يحوي العناصر الكيماوية مرتبة حسب أوزانها الذرية

أجندة

- 4 جدول بأيام السنة مرتب حسب تعاقبها الزمني
- 5 سجل المواعيد وأحداث وزمن حدوثها

جدول

- 6 مصفوفة بيانات مكونة من صفوف وأعمدة
- 7 نهر صغير
- 8 رتب الأشياء على شكل جدول

نهر النيل

- 9 نهر في إفريقيا ينبع من بحيرة فيكتوريا وبحيرة تانا ويصب في البحر الأبيض المتوسط ماراً في السودان ومصر.

مصفوفة

- 10 ترتيب لأشياء جنباً الى جنب

ترتيب

- 14 تنظيم الأشياء بصورة ممنهجة
- 15 موضع الشيء ضمن نسق تعاقبي

Gloss: describes a concept

تيار مائي كبير يتدفق بمجرى

الأخدود الذي يتدفق فيه تيار مائي كبير

زجر بصوت قوي

زجر

- 16 عنف بنية النهي أو الطرد أو التوبيخ

# What/ Why a Gloss

according to strict ontological guidelines[J06]

التعريفات

A **gloss**: is an auxiliary *informal (but controlled) account* of the intended meaning of a linguistic term, for the commonsense perception of humans.

صف

1 مجموع من التلاميذ في مستوى تعليمي واحد

2 ترتيب لأشياء تكون فيه مستوية وجنباً الى جنب

الجدول الدوري

3 جدول يحوي العناصر الكيماوية مرتبة حسب أوزانها الذرية

أجندة

4 جدول بأيام السنة مرتب حسب تعاقبها الزمني

5 سجل المواعيد وأحداث وزمن حدوثها

جدول

6 مصفوفة بيانات مكونة من صفوف وأعمدة

7 نهر صغير

8 رتب الأشياء على شكل جدول

نهر النيل

9 نهر في إفريقيا ينبع من بحيرة فيكتوريا وبحيرة تانا ويصب في البحر الأبيض المتوسط ماراً في السودان ومصر.

مصفوفة

10 ترتيب لأشياء جنباً الى جنب

نهر

11 تيار مائي كبير يتدفق بمجرى

12 الأخدود الذي يتدفق فيه تيار مائي كبير

13 زجر بصوت قوي

ترتيب

14 تنظيم الأشياء بصورة ممنهجة

15 موضع الشيء ضمن نسق تعاقبي

زجر

16 عنف بنية النهي أو الطرد أو التوبيخ

A gloss is supposed to render factual knowledge that is critical to understand a concept, but that e.g. is implausible, unreasonable, or very difficult to formalize and/or articulate explicitly. (NOT) to catalogue general information and comments, as e.g. conventional dictionaries and encyclopedias usually do, or as `<rdfs:comment>`.

# Arabic Ontology: Gloss Guidelines

قواعد كتابة التعريفات

## What should and what should not be provided in a gloss:

1. Start with the *principal/super type* of the concept being defined.

E.g. 'Search engine': "A computer program that ...", 'Invoice': "A business document that...",  
'University': "An institution of ...".

الجنس الأعلى للمفهوم المراد تعريفة

2. Focus on distinguishing characteristics and intrinsic prosperities that differentiate the concept out of other concepts.

E.g. Compare, 'Laptop computer':

"A computer that is designed to do pretty much anything a desktop computer can do, it runs for a short time (usually two to five hours) on batteries".

تليها الصفات الجوهرية /

"A portable computer small enough to use in your lap...".

التميزة للمفهوم المراد

3. Written in a form of propositions, offering the reader inferential knowledge that help him to construct the image of the concept.

E.g. Compare 'Search engine':

"A computer program for searching the internet, it can be defined as one of the most useful aspects of the World Wide Web. Some of the major ones are Google, ....";

A computer program that enables users to search and retrieves documents or data from a database or from a computer network..."

ضاي، بطريقة تقود القارئ لاستنباط المعنى



# Arabic Ontology: **Gloss Guidelines**

4. Use supportive examples : إستخدم الامثلة مسموح ولكن  
بتحفظ شديدا وحالات معينة
- To clarify cases that are commonly known to be false but they are true, or that are known to be true but they are false;
  - To strengthen and illustrate distinguishing characteristics (e.g. define by examples, counter-examples).

Examples can be types and/or instances of the concept being defined.

5. Be consistent with formal definitions/axioms.

6. Be sufficient, clear, and easy to understand.

➔ **WordNet glosses do not follow such ontological guidelines**

# Arabic Ontology: Gloss Guidelines

As a gloss starts with a supertype of concept being defined, try to read the gloss as the following, to verify what you do is correct:

- جدول: مصفوفة بيانات مكونة من صفوف وأعمدة.**  
**جدول: ترتيب بيانات جنباً الى جنب على شكل صفوف وأعمدة.**  
**جدول: تنظيم بيانات بصورة ممنهجة جنباً الى جنب على شكل صفوف وأعمدة.**

صف

1 مجموع من التلاميذ في مستوى تعليمي واحد

2 ترتيب لأشياء تكون فيه مستوية وجنباً الى جنب

الجدول الدوري

3 جدول يحوي العناصر الكيماوية مرتبة حسب أوزانها الذرية

أجندة

4 جدول بأيام السنة مرتب حسب تعاقبها الزمني

5 سجل المواعيد وأحداث ووزمن حدوثها

جدول

6 مصفوفة بيانات مكونة من صفوف وأعمدة

7 نهر صغير

8 رتب الأشياء على شكل جدول

نهر النيل

9 نهر في إفريقيا ينبع من بحيرة فيكتوريا وبحيرة تانا ويصب في البحر الأبيض المتوسط ماراً في السودان ومصر.

مصفوفة

10 ترتيب لأشياء جنباً الى جنب

نهر

11 تيار مائي كبير يتدفق بمجرى

12 الأخدود الذي يتدفق فيه تيار مائي كبير

13 زجر بصوت قوي

ترتيب

14 تنظيم الأشياء بصورة ممنهجة

15 موضع الشيء ضمن نسق تعاقبي

زجر

16 عنف بنية النهي أو الطرد أو التوبيخ

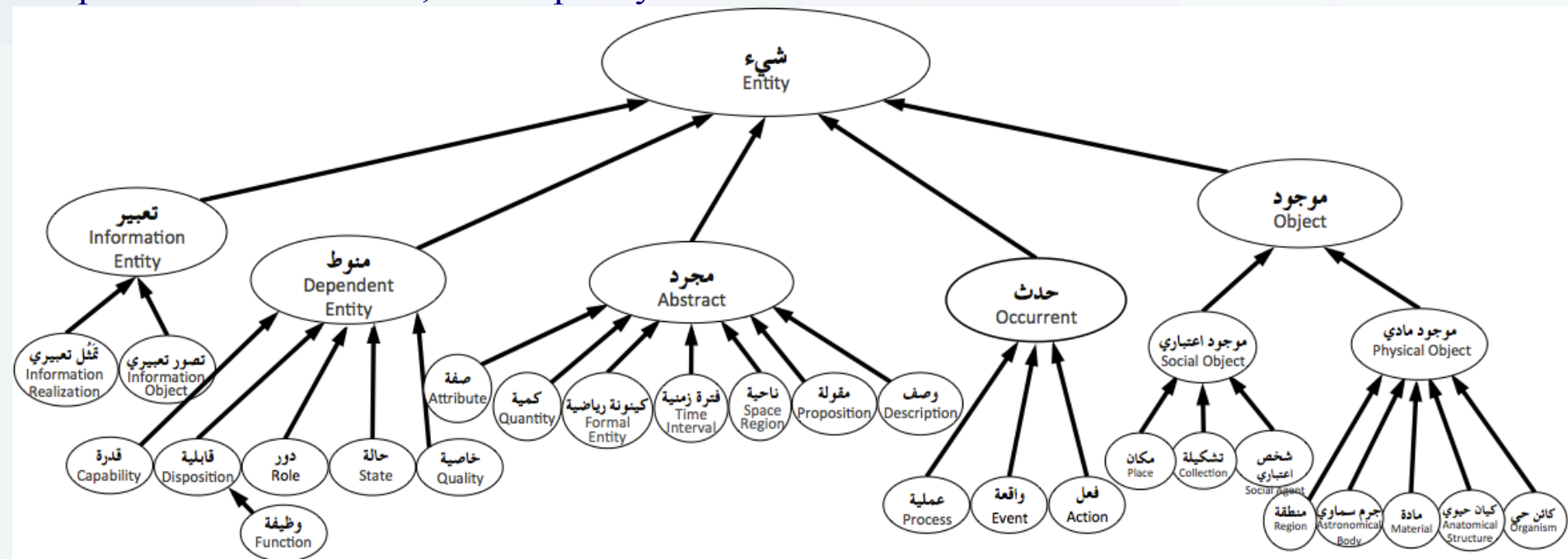
- Part 1: The Arabic Ontology Design
- Part 2: Gloss in the Arabic Ontology
-  • Part 3: The Top Levels of the Arabic Ontology
- Part 4: Arabic Ontology Vs WordNet
- Part 5: Building Synsets Automatically

# The Core (Top Levels) of the Arabic Ontology

الحدود العلوي: هي أمهات المعاني لجميع الكلمات العربية

Arabic Core Ontology: the top levels of the Arabic Ontology - built manually, and carefully considering the philosophical and historical aspects of the Arabic concepts/terms, as well as BFO and DOLCE upper level ontologies. Different from the 25 unique beginners in WordNet

Top 3 levels shown here, for simplicity



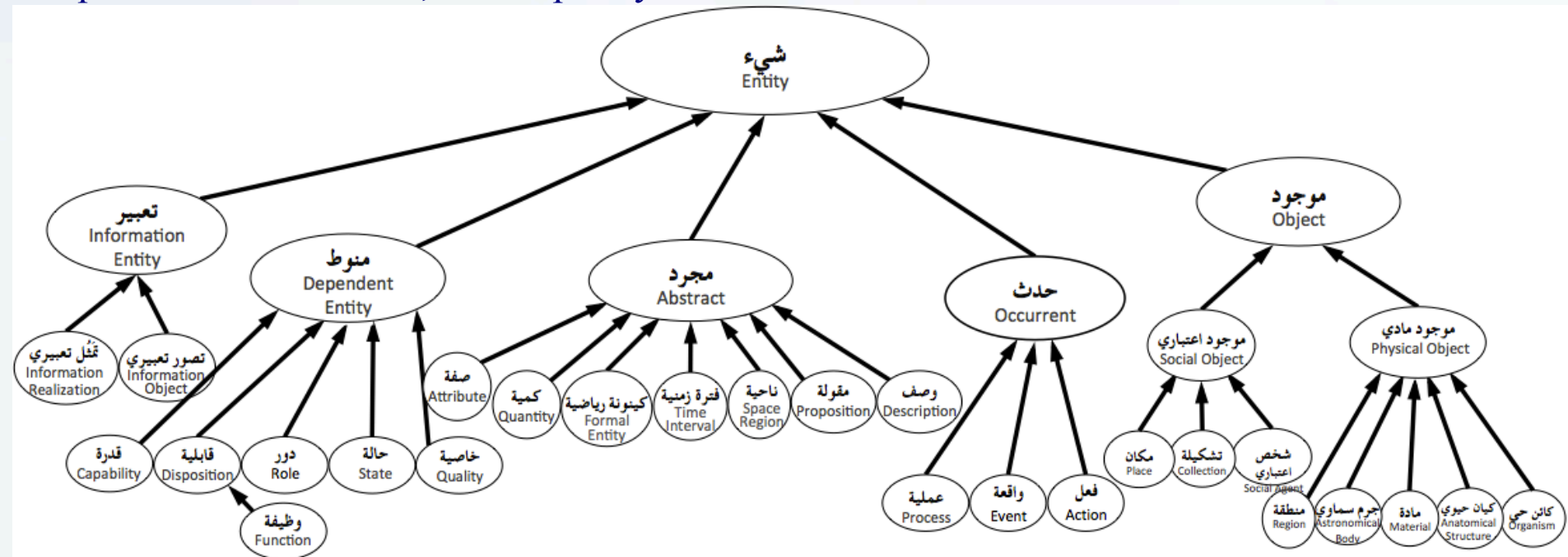
# Why these Core Concepts (Top Levels)?

الحدود العلوي: هي أمهات المعاني لجميع الكلمات العربية

Why this these top levels are so important:

- The 10<sup>th</sup> level of this core ontology should top all Arabic concepts and levels.
- This allow us to detect any problems in the tree/relations!
- The core Ontology governs the correctness and the evolution of the whole Arabic Ontology.

Top 3 levels shown here, for simplicity



- Part 1: The Arabic Ontology Design
- Part 2: Gloss in the Arabic Ontology
- Part 3: The Top Levels of the Arabic Ontology
-  • Part 4: Arabic Ontology Vs WordNet
- Part 5: Building Synsets Automatically

# Ontology Versus WordNet

## Meaning (called Ontological Precision):

WordNet: based on what native speakers agree roughly

Ontology: based on Scientific and philosophical findings.

## Classification:

WordNet: based on what native speakers agree roughly

e.g., (Student IsA person) (Nile IsA River)

Ontology: based on strict formal methodologies

e.g., (Student IsA role) (Nile InstanceOf River)

## Formal Specification:

WordNet: logically vague

Ontology: strictly formal

→ WordNet as a linguistic ontology, though it needs lots of cleaning!

→ Linguistic ontologies are difficult to build but they are immune to changes

# Arabic Ontology Vs WordNet

Unlike WordNet, the Arabic Ontology is:

## 1. Philosophically well founded:

- Focuses on intrinsic properties;
- All types are rigid;
- The top level is derived from known Top Level Ontologies.

## 2. Strictly formal:

- Semantic relations are well-defined relations.

## 3. Strictly-controlled glosses

- The content and structure of the glosses is strictly based on ontological principles.

- ➔ The Arabic Ontology can be used as an Arabic WordNet, but is more.
- ➔ The Arabic Ontology follows a similar structure to WordNet, and is well-mapped.



- Part 1: The Arabic Ontology Design
- Part 2: Gloss in the Arabic Ontology
- Part 3: The Top Levels of the Arabic Ontology
- Part 4: Arabic Ontology Vs WordNet
-  Part 5: Building Synsets Automatically

(this is a research part)





















