

The Arabic Ontology Basics

الأنطولوجيا العربية

Mustafa Jarrar

[Birzeit University, Palestine](#)

mjarrar@birzeit.edu

www.jarrar.info



Watch this lecture and download the slides



Course Page: <http://www.jarrar.info/courses/ArabicOntology.pdf>

More Online Courses at: <http://www.jarrar.info/courses/>

The lecture is based on:

Please use these articles if when citing the Arabic Ontology

- Mustafa Jarrar, Hamzeh Amayreh: Linguistic Search Engine. Proceedings of the Web conference (WWW2019), ACM, 2019
- Mustafa Jarrar: The Arabic Ontology - An Arabic Wordnet with Ontologically Clean Content. Applied Ontology Journal. IOS, 2019.
- 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>
- 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>

The Arabic Ontology



Part 1: Motivation - Why the Arabic Ontology

- Part 2: The Arabic Ontology Design
- Part 3: The Top Levels of the Arabic Ontology
- Part 4: Ontology Benchmarking Methodology
- Part 5: Glosses in the Arabic Ontology

Where the Arabic Ontology can be used?

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.**

Application Ontology vs. Linguistic Ontology

The importance of linguistic ontologies is growing rapidly.

Application ontology

To represent the semantics of a certain domain/application, e.g., the Gene Ontology, the FOAF ontology, the Palestinian e-government ontology.

- Each term convey one concept (no polysemy).
- Represents (/Benchmarked to) application's knowledge.
- Used only by a certain application, or a class of applications.

Linguistic ontology

To represent the semantics of terms in a human language, independently of a particular application.

- Each word may convey several concepts (Polysemy).
- Represents common-sense knowledge (/State-of-art scientific discoveries).
- Can be used for general purposes.

WordNet Vs Arabic Ontology

- The Arabic Ontology is:
 - a next generation of WordNet.
 - an ontologically-clean Arabic WordNet
 - follows the same structure (/data model) as wordnet
 - fully mapped to Wordnet.
- But there are critical foundational differences.

WordNet Vs Arabic Ontology

The main foundational differences:

- benchmarked on native-speaker's naïve knowledge;
 - No top levels;
 - Concepts are informal; concepts without individuals (e.g., horizon) are allowed.
 - Glosses are example contexts.
- benchmarked on state-of-art scientific discoveries
 - Governed by scientifically & philosophically well-established top levels;
 - Concepts are formal; a concept is a set of individuals. (e.g., horizon is not allowed.
 - Glosses are strictly formulated, and focus on the distinguishing characteristics.

The Arabic Ontology

- Part 1: Motivation - Why the Arabic Ontology



Part 2: The Arabic Ontology Design

- Part 3: The Top Levels of the Arabic Ontology
- Part 4: Ontology Benchmarking Methodology
- Part 5: Glosses in the Arabic Ontology

The Arabic Ontology

- A project started in 2010, at Birzeit University, Palestine.
- Tree of the Arabic concepts -Classification of the meanings of the Arabic terms, based on state-of-art science, rather than on speakers' naïve knowledge

<http://ontology.birzeit.edu>

The screenshot displays the website interface for the Arabic Ontology project. At the top, there is a search bar containing the word 'time'. To the right of the search bar, there are language options for 'عربي' (Arabic) and 'English', and a user profile icon. Below the search bar, the page shows search results for 'time'. The results are organized into a list of entries, each with a title in Arabic and English, a brief description, and a 'TypeOf' label. The entries include:

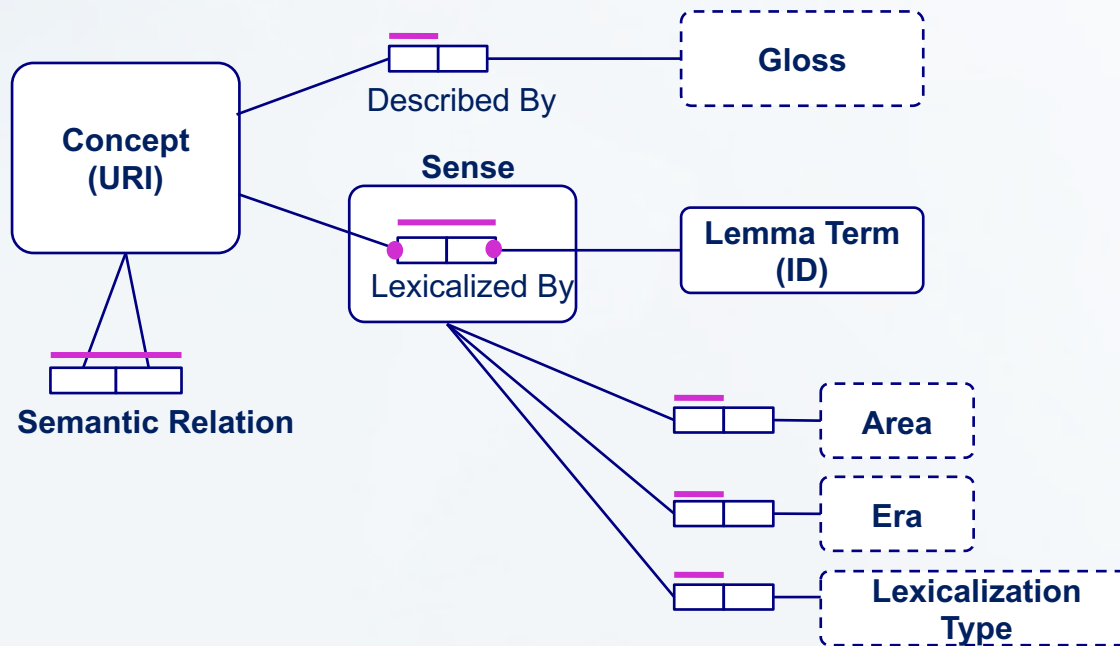
- time** (زمن): وسط متجانس غير محدود يمر فيه الأحداث متلاحقة ، والمدّة جزء منه . وقد يطلق على مدّة معينة. (Philosophy Lexicon ©)
- time** (زمن): هذه أرسطو مفاهيم الحركة ، ولفظ يبيّن وبين المكان ، وما قامت الحركة متصلة للزمن متصل . (Philosophy Lexicon ©)
- time** (زمن): هذه كلفظ . الزمن صورة أولية ترجع إلى الحسابية الداخلية بصفة مباشرة وإلى الحسابية الخارجية بصفة غير مباشرة ، لكل إنسان إما هو حدس نفسي له موقعه من الزمان##2# هذه أرسطو مفاهيم الحركة ، المزيد... (Philosophy Lexicon ©)
- time** (زمن): هذه برجماتيون تغير مستمر متصل يصحح معه الحاضر ماضياً، وبأني الذاهن أن يضل في التيار الزمني ، بل يجمع كل ما يتصل به ويربطه بفكرة معينة فالزمن مرتبط عنده بالديمومة##3# أرسطو مفاهيم غير محتمر المزيد... (Philosophy Lexicon ©)
- الزمن** (Time): 1 - الزمن الوقت كثيره وقبله. وهو المدّة الواقعيّة بين حادثتين أو لاهما سابقة وتاليهما لاحقة، ومدّه زمن الحوادث، وزمان الكليات، وزمان الجاهلية. وجمع الزمان الأزمنة، لقولنا: السنة أربعة أزم للزمن... (Philosophy Lexicon P1 and P2 ©)

On the right side of the page, there is a section titled 'ONTOLOGY' which shows a search result for 'time' with 4 results. The first result is:

- Time** (زمن) | **قَرَأَ** (Time): Abstract entity representing a region in the timeline, realized by its starting and ending points, its length represents the temporal dimension of events or objects. معيّنه يمكن جزءاً من خطّ الزمن المتحرك، تتركه ذاته بلفظة بداية ونقطة نهاية طولها يمكن زمن أحداث أو موجودات مثال : يستغرق دوران الأرض حول الشمس زمناً يعرف بالسنة (TypeOf: (Occurent) 293570)
- Interval** | **Time Interval** (قَرَأَ | قَرَأَ زَمَنَةً) (Interval): Amount of time, its length is calculated based on the the temporal dimension of astroligical events, its starting and ending points are not equal, and has no gaps. زمن يُكتمر ويُحسب بناءً على ما يلفظ به من أحداث فلكية، له بداية ونهاية غير متساوية، تظفر من التغيرات الزمنية مثال : الليل هو فترة زمنية بين غروب الشمس وشرورها (TypeOf: (Time) 293572)
- Duration** | **Time Duration** (قَرَأَ | حِينٌ | مَدَّةٌ زَمَنِيَّةٌ زَمَانِيَّةٌ | مَدَى | أَمَدٌ | أَمَلٌ) (Duration): Amount of time, its length depends on the life (i.e., the temporal dimension) of events or objects.

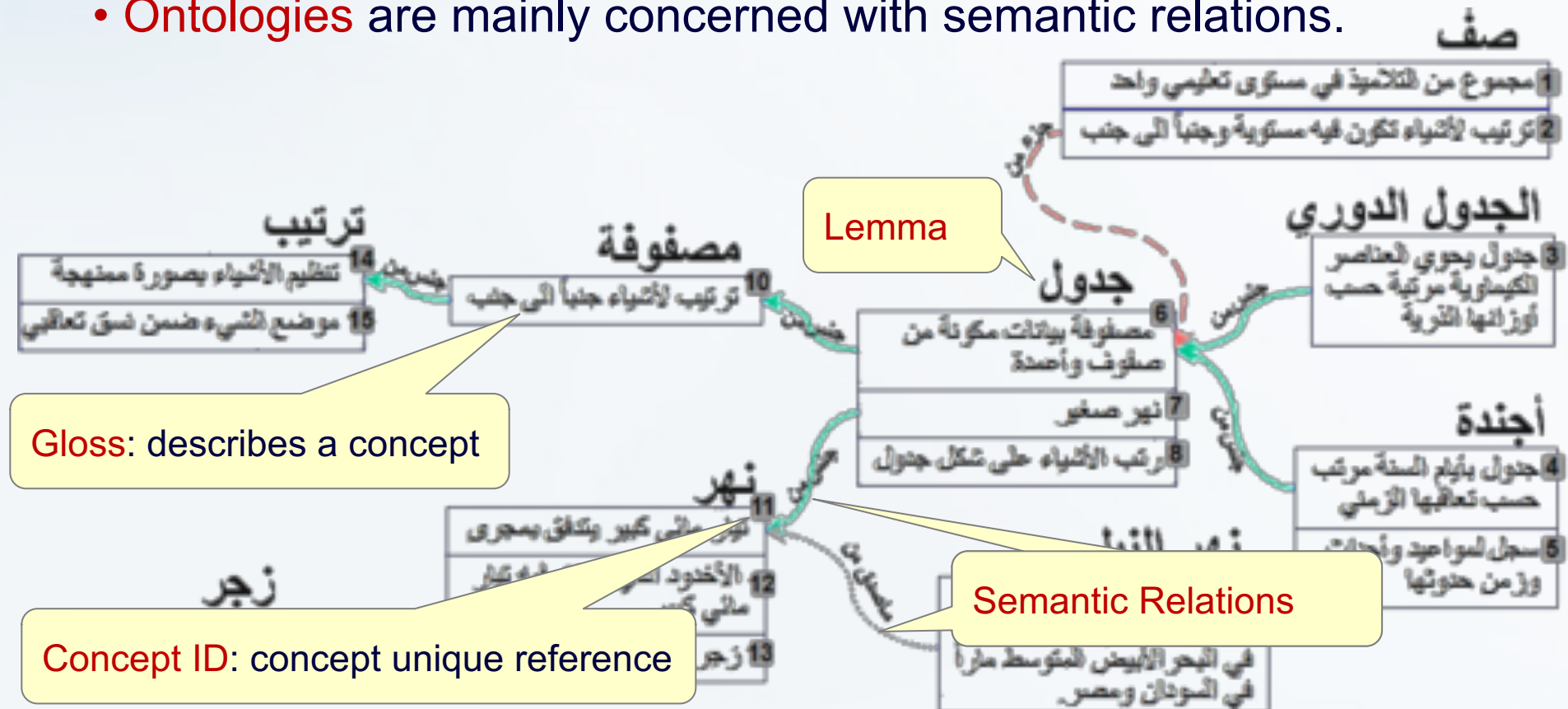
Arabic Ontology: Data Model (Simplified)

The general structure (i.e. core data model) of the Arabic Ontology is similar to the structure of WordNet – So to help in concept-mapping and interoperability.



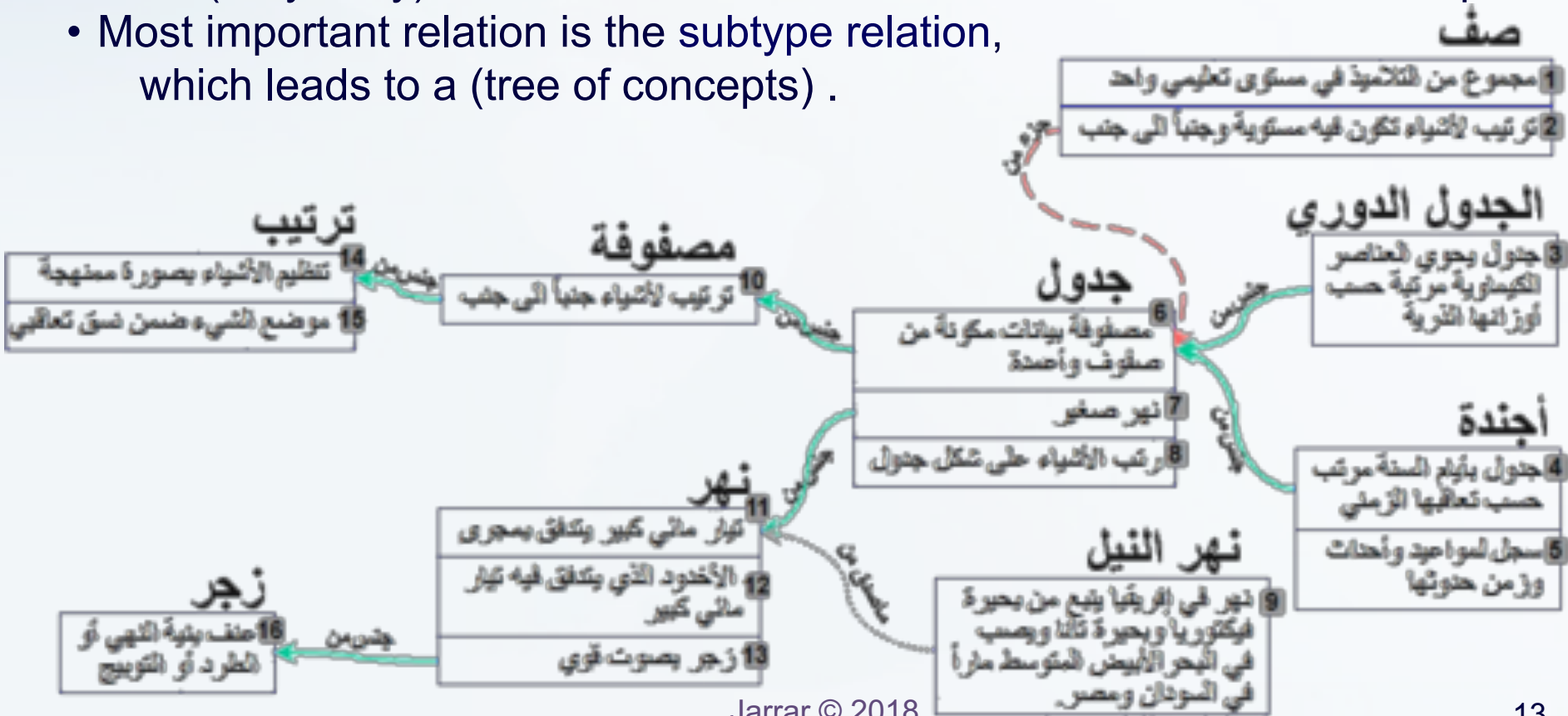
Lexical vs. Semantic Relationships

- **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.

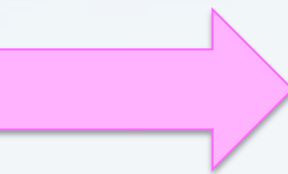


Arabic Ontology

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



The Arabic Ontology

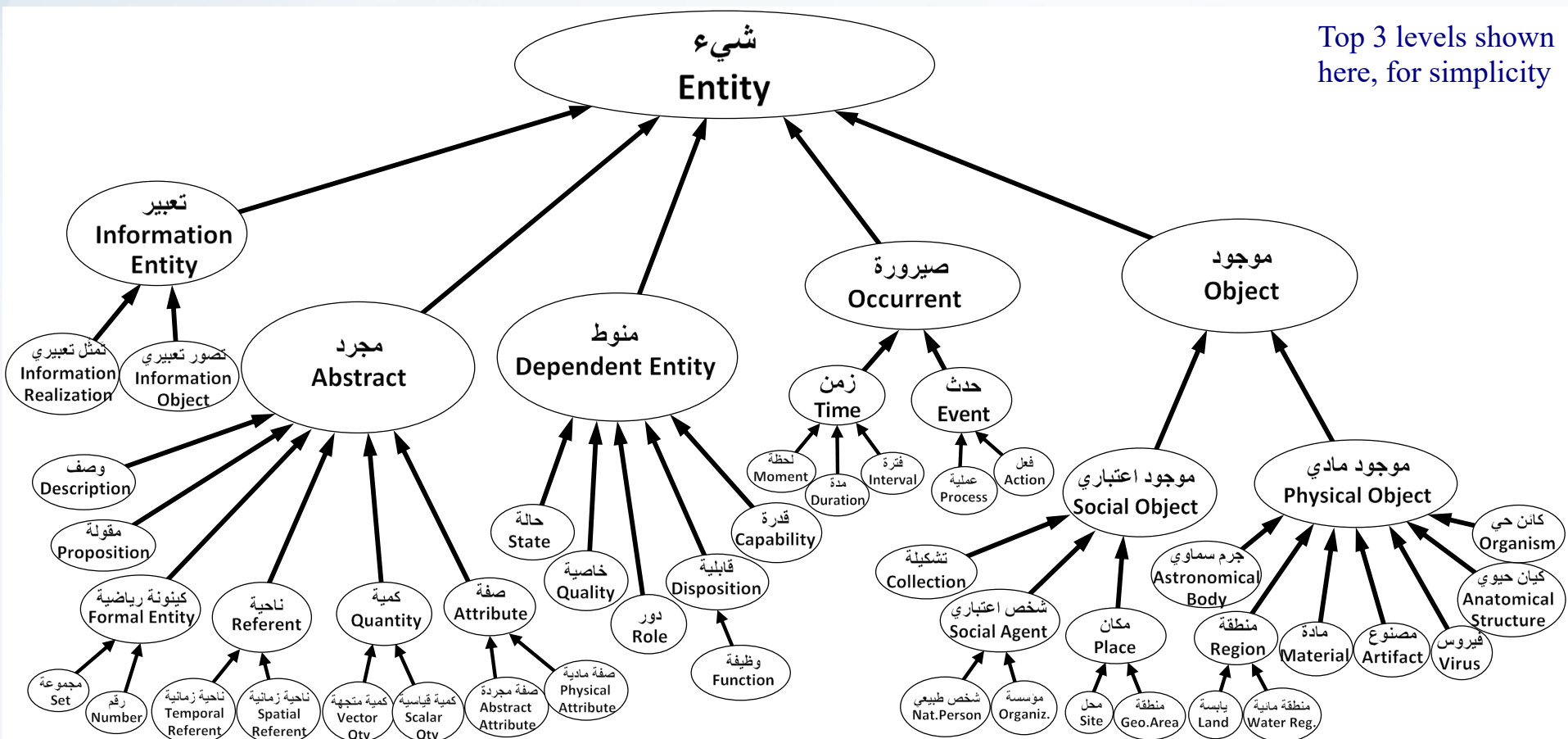
- Part 1: Motivation - Why the Arabic Ontology
- Part 2: The Arabic Ontology Design
-  Part 3: The Top Levels of the Arabic Ontology
- Part 4: Ontology Benchmarking Methodology
- Part 5: Glosses in the Arabic Ontology

The Ontology Top Levels

الحدود العليا: هي أمهات المعاني لجميع الكلمات العربية

The top levels of the Arabic Ontology tree are the most abstract concepts in Arabic; they are philosophically and logically well defined, and linked with BFO and DOLCE upper level ontologies.

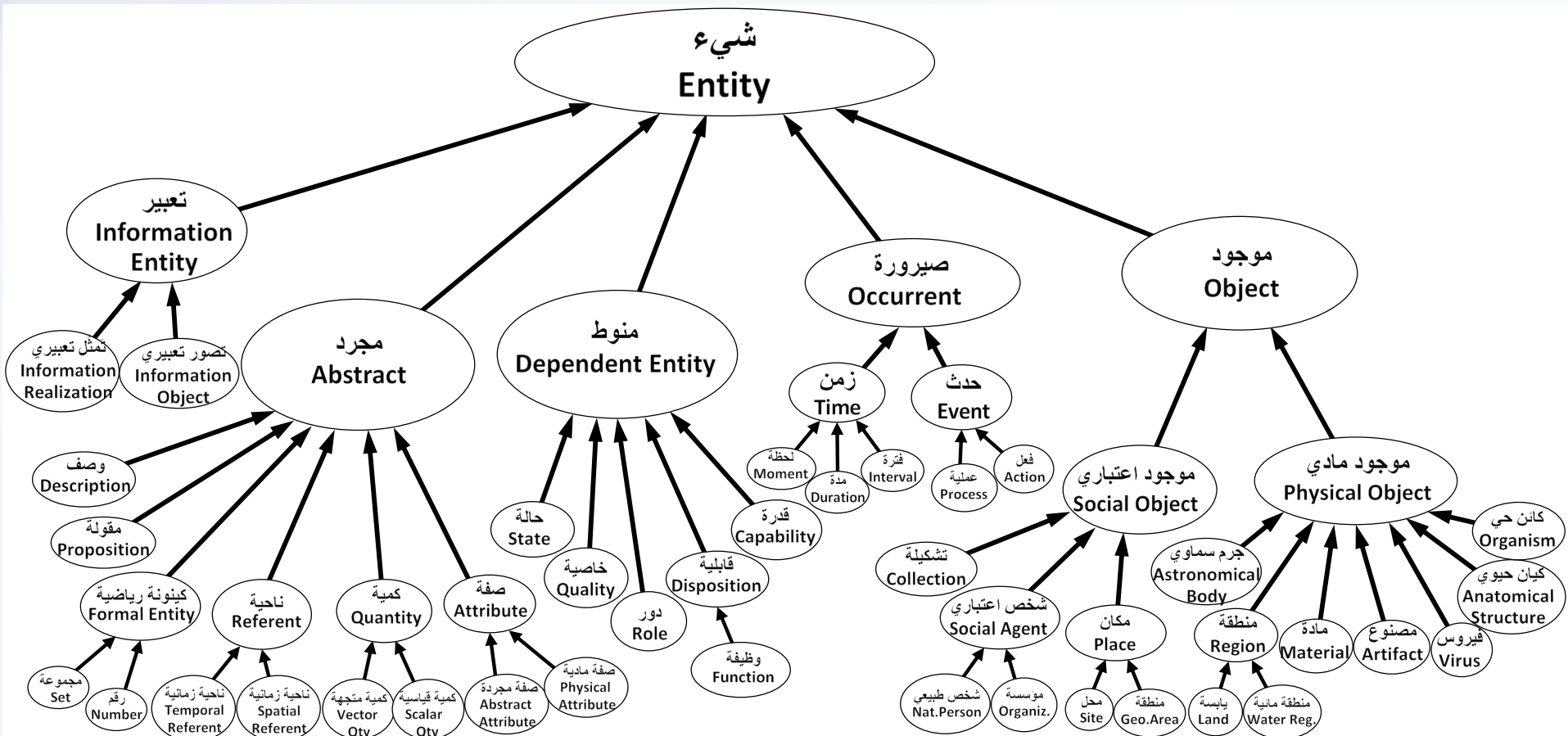
Top 3 levels shown here, for simplicity



The Ontology Top Levels

Why these top levels are so important:

- Derive/specialize all Arabic concepts from these top levels.
- Allows us to detect any logical and ontological mistakes in the lower levels.
- Used to governs the correctness and the evolution of the lower levels.



The Arabic Ontology

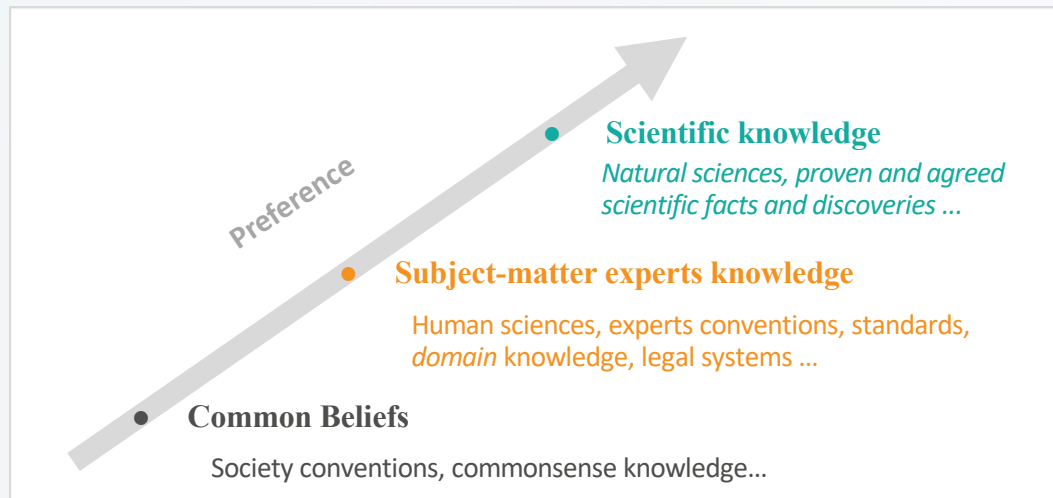
- Part 1: Motivation - Why the Arabic Ontology
- Part 2: The Arabic Ontology Design
- Part 3: The Top Levels of the Arabic Ontology
- Part 4: Ontology Benchmarking Methodology
- Part 5: Glosses in the Arabic Ontology



Benchmarking Methodology

The ontological precision of the classifications benchmarked against the following:

- **Scientific knowledge**, which scientists typically accept on the basis of experimentation and verification and commonly agree about. If no mature answer is found in the state-of-art *scientific discoveries*, then against,
- **Subject-matter experts' and domain knowledge and conventions**. If no satisfactory answer can be synthesized from *experts' knowledge*, then against,
- **Common-sense knowledge** that can be repeatedly found in quality lexicons and literature.



The Arabic Ontology

- Part 1: Motivation - Why the Arabic Ontology
- Part 2: The Arabic Ontology Design
- Part 3: The Top Levels of the Arabic Ontology
- Part 4: Ontology Benchmarking Methodology
- Part 5: Glosses in the Arabic Ontology

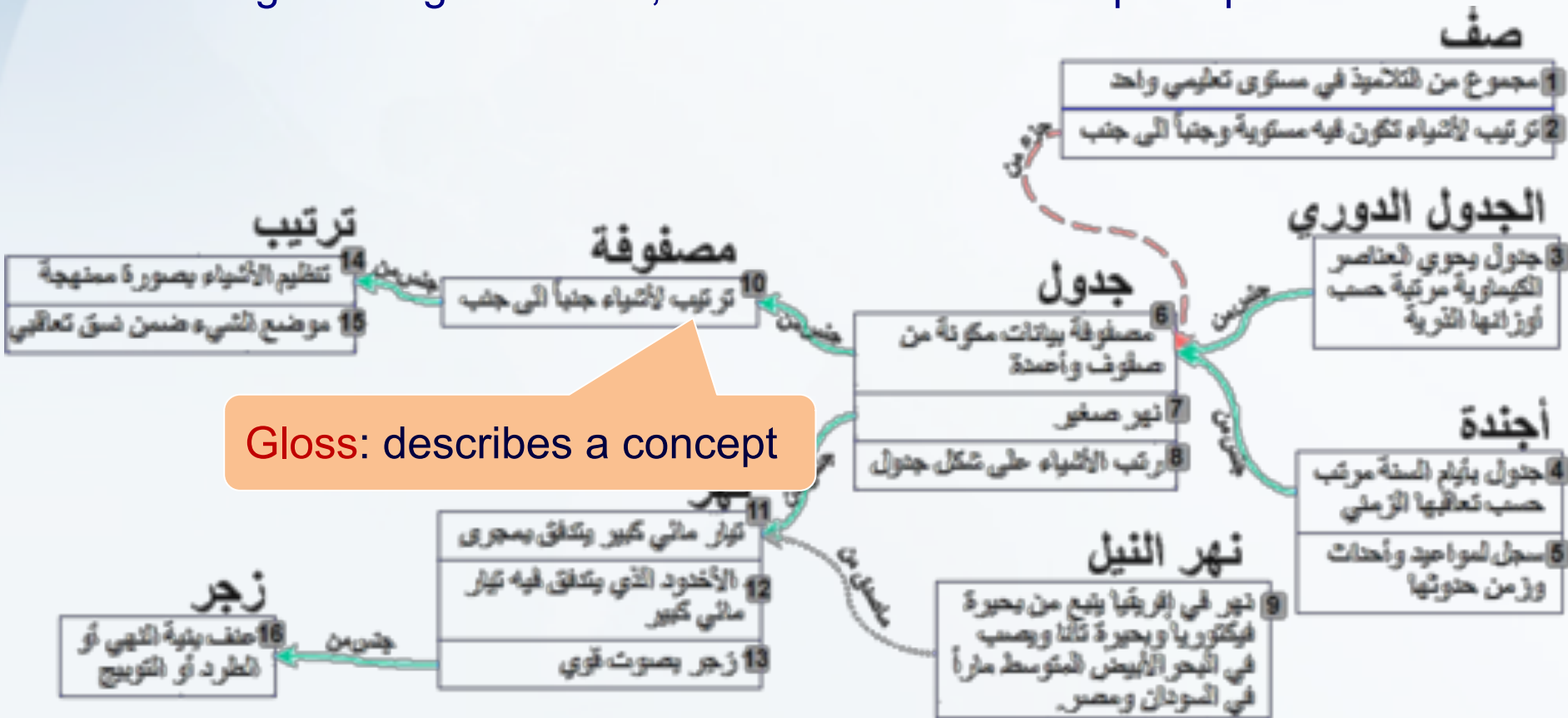


What/ Why a Gloss

كتابة التعريفات

according to strict ontological guidelines[J06]

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



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 ...".
يبدأ التعريف بالجنس الاعلى للمفهوم المراد تعريفه

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/المميزة

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

للمفهوم المراد تعريفه

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...".

يكتب التعريف على شكل قضايا، بطريقة تقود القارئ لاستنباط المعنى

Arabic Ontology: **Gloss Guidelines**

4. Use supportive examples :

استخدام الامثلة مسموح ولكن بتحفظ
شديد 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 مسكوية وجنباً الى جنب

الجدول الدوري

- 3 جدول يحوي العناصر الكيميائية مرتبة حسب أوزانها الذرية

جدول

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

مصفوفة

- 10 ترتيب لأشياء جنباً الى جنب

ترتيب

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

أجندة

- 4 جدول بأيام السنة مرتب حسب تعاقبها الزمني
- 5 سجل للموايد وأحداث ووزن محتوياتها

نهر

- 11 نوار مائي كبير يتدفق بمجرى
- 12 الأخود الذي يتدفق فيه تيار مائي كبير
- 13 زجر بصوت قوي

نهر النيل

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

زجر

- 16 صنف بثنية النهي أو الطرد أو التوبيج

Gloss: describes a concept

References

1. Mustafa Jarrar, Hamzeh Amayreh: Linguistic Search Engine. Proceedings of the Web conference (WWW2019), ACM, 2019
2. Mustafa Jarrar: The Arabic Ontology - An Arabic Wordnet with Ontologically Clean Content. Applied Ontology Journal. IOS, 2019.
3. Mustafa Jarrar, Fadi Zaraket, Rami Asia, Hamzeh Amayreh: **Diacritic-Based Matching of Arabic Words**. ACM Transactions on Asian Language Information Processing. (Forthcoming)
4. Mustafa Jarrar, Werner Ceusters: **Classifying Processes and Basic Formal Ontology**. Proceedings of the 8th International Conference on Biomedical Ontology (ICBO 2017), Newcastle, UK. 2017
5. Mustafa Jarrar: **Building a Formal Arabic Ontology (Invited Paper)**. In proceedings of the Experts Meeting on Arabic Ontologies and Semantic Networks. Alecco, Arab League. Tunis, July 26-28, 2011.
6. 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.
7. Mustafa Jarrar, Anton Deik: **The Graph Signature: A Scalable Query Optimization Index for RDF Graph Databases Using Bisimulation and Trace Equivalence Summarization**. International Journal on Semantic Web and Information Systems, 11(2), 36-65, DOI: 10.4018/IJSWIS.2015040102. April-June 2015
8. Mustafa Jarrar: Lecture Notes on Lexical Semantics and Multilingualism. Birzeit University, Palestine. 2018
9. Mustafa Jarrar: Lecture Notes on Introduction to Wordnets. Birzeit University, Palestine. 2018
10. Mustafa Jarrar: Lecture Notes on the Arabic Ontology Basics. Birzeit University, Palestine. 2018
11. Mustafa Jarrar, Anton Deik, Bilal Faraj: **Ontology-based Data and Process Governance Framework -The Case of e-Government Interoperability in Palestine**. Proceedings of the IFIP International Symposium on Data-Driven Process Discovery and Analysis (SIMPDA'11). Pages(83-98). ISBN 978-88-903120-2-1. Campione, Italy. June, 2011.
12. Mustafa Jarrar and Robert Meersman: **Ontology Engineering -The DOGMA Approach**. Book Chapter in "Advances in Web Semantics I". Chapter 3. Pages 7-34. LNCS 4891, Springer.ISBN:978-3540897835. (2008).
13. Mustafa Jarrar: **Mapping ORM into the SHOIN/OWL Description Logic- Towards a Methodological and Expressive Graphical Notation for Ontology Engineering**. In OTM 2007 workshops: Proceedings of the International Workshop on Object-Role Modeling (ORM'07). Pages (729-741), LNCS 4805, Springer. ISBN: 9783540768890. Portugal. November, 2007
14. Mustafa Jarrar, Maria Keet, and Paolo Dongilli: **Multilingual verbalization of ORM conceptual models and axiomatized ontologies**. Technical report. STARLab, Vrije Universiteit Brussel, February 2006.