مجلة الآداب والعلوم الاجتماعية Journal of letters and Social Sciences (JLSS)
https://www.asjp.cerist.dz/en/PresentationRevue/4

نحو هندسة آليّة لبناء الأنطولوجيات العربيّة ومحركات البحث المعجميّة - جامعة بيزيت فلسطين أنموذجا-

Toward an automated architecture for building Arabic ontologies and lexical search engines

- Bzeit Palestine University as an example-

د. وسيلة داودي
Dr. Wassila Daoudi
جامعة محمّد لمين دباغين سطيف2، الجزائر
wassiladaoudi@yahoo.com

معلومات حول المقال
تاريخ الاستلام 31-12-2024
تاريخ القبول 01-10-2025

الكلمات المفتاحية
الأنطولوجيا
الدلالة المعجمية
اللسانيات الحاسوبية
الحاسوب
المعجمات الإلكترونية

ملخص

من بين التّحدّيات الحضاريّة والمسؤوليّات التّاريخيّة ما تواجهه لغتنا العربيّة في استكمال اللسانيّات الحاسوبيّة وتطبيقها، وهي بذلك تتجهّز لنقلة مجتمعيّة نوعيّة تستلزم من أصحاب هذه اللغة الكشف عن خصائصها حتّى يسهل معالجتها آليّا، وهو ما قامت به أغلب الدّول مع لغاتها ولهجاتها، فأصبح لزاما علينا تطوير لغتنا من خلال أساليب الذكاء الاصطناعيّ والقدرة الآليّة على التّعامل مع اللغة. ولمّا كانت المعالجة الآليّة لا يمكنها إلا أن تتعامل مع ما هو واضح ومحدّد، فإنّ هذا الأمر يشكّل عائقا ومنعطفا مهمّا في مستقبل اللغة العربيّة، الأمر ذاته الّذي جعل خصوصيتها تواجه إشكالا على مستوى الدلالة الّذي يصعب ضبطه، فغدا بذلك إشكالا تواجهه قواعد البيانات في سبيل تحقيق رقمنة ناجحة بكل المقاييس، وعليه نقدم هذه الورقة البحثية للوقوف على هذه القضية من خلال تجربة «جامعة بيرزيت الفلسطينية» وفق دراسة وصفية. وقد حاولنا من خلال هذه الدراسة الّتي اعتمدنا فيها على جديد التكنولوجيا وتوظيفها في حوسبة اللغة العربية، مبرزين دور المعجمات الإلكترونية في سبيل نمذجة المعرفة وتمثيلها بشكل منهجيّ سليم يسهّل على الحاسوب معالجته، وسعيا لتحقيق ذلك فإنّ بناء الأنطولوجيّات العربية هي الأساس للتّبادل السليم والفعّال للبيانات، من خلال تحديد دقيق للمعاني الدّلاليّة المراد تبادلها.

مقدمة

انصهرت الثورة التكنولوجيّة بصورة متزايدة في كيان المجتمعات الإنسانيّة بشكل ملفت للنّظر، وطفت بذلك قضيّة المعالجة الآليّة للغة العربيّة والقدرة على تخزين ومعالجة البيانات، طالما نبّه الأخصائيّون إلى « التبسيط المسرف للفنيّين من جانب، والتّجاهل الحاد من قبل اللغويّين، من جانب آخر[...] وهي أيضا من الأهميّة والحيويّة بصورة تفرض أقصى درجات الاهتمام، وتجعلها في مقدّمة الأولويّات لإعداد مجتمعاتنا العربيّة لمجتمع الغد القريب: مجتمع المعلومات الّذي ستسوده عمالة المعلومات وصناعاتها وخدماتها» (علي، 1988)، وعلى تعدّد تسميات هذه المرحلة وهول الفوارق الّتي ستحدثها واتّساع الهوة الفاصلة بين العالم النّامي والعالم الصّناعي المتطوّر، فإنّه من الواجب إدراك مدى خطورتها، والسّعي الحثيث لتقليص آثارها واجب ملزم اتّجاه الحاضر واتّجاه الأجيال القادمة. (علي، 1988) ميدان الدّراسات والبحوث العلميّة في اللسانيّات الحاسوبيّة هو ميدان علميّ وتطبيقيّ واسع جدّا؛ فهو يشمل تطبيقات كثيرة كالتّرجمة الآليّة والإصلاح الآلي للأخطاء المطبعيّة وتعليم اللّغات بالحاسوب والعمل الوثائقيّ الآلي، وتطبيق الآلات بالتركيب الاصطناعي للأصوات اللغويّة وغيره كثير. (صالح، 2012)

يرى كبار الفنيّين أنّ المعماريّة المقترحة لحاسوب الجيل الخامس ونظم تشغيلها تُؤكّد لنا الدّور الرّئيسي الّذي ستلعبه اللغة في عالم الحواسيب الذكيّة، وستواجه الدّول

المتخلّفة تقنيّا الكثير من التّحدّيات القاسية، ولذلك عليها استيعاب الأبعاد اللغويّة التقنيّة لمشروع الجيل الخامس، بل حتّى ما يتعلّق بمجالات التعليم والتثقيف والإعلام والإدارة والتّخطيط والتّصنيع. (علي، 1988)

وعلى هذا الأساس، فإنّ «حوسبة المعجم العربي في الوقت الرّاهن أضحى مطلبا ملحّا أكثر من أيّ وقت مضى، من أجل سدّ الفجوة المعجميّة بين العربيّة واللغات المتقدّمة تقانيّا ومعرفيّا» (مهديوي، 2012)، ولم يعد ذلك من قبيل الرّفاهيّة الفنيّة بل واجبا أساسيّا تفرضه اعتبارات عدّة منها طبيعة مضمونه، وتنظيمه، وتحديثه، وخدماته للبشر ولنظمهم الآليّة، إضافة إلى طبيعة العلم باعتباره معرفة ومنهجا. (علي، 1988)

وعلى أساس ما سبق ذكره فقد نبّه كبار اللّسانيين إلى ضرورة الحوار بين الباحثين المختلفي التكوين والميادين بتبادل الأفكار والنقاشات والفهم والانتقادات والاقتراحات بلغة التخصص الّتي ينتمي إليها، ويستدعي ذلك أيضا بتلقي دروس ضمن التكوين العلمي للرّياضيين والمهندسين في المعلوميّات وحاملي الليسانس في اللغة العربيّة وغيرهم. (صالح، 2012)

1-المعالجة الآليّة للمستوى الدلالي للغة العربيّة

يعود الانشغال بالدّراسات الدلاليّة للّغة العربيّة بداية من القرون الثالث والرّابع والخامس الهجريّة إلى سائر القرون التالية لها، ولا يعبّر هذا التّأريخ المبكّر إلا على النّضج الّذي أحرزته العربيّة وأصّله الدارسون، غايتهم في ذلك تأسيس علم عربيّ له شخصيّته الّتي تساعد على إنجاز تطبيقات حديثة بوضوح ووعي لدى اللغويّين والنقّاد. (الداية، 1996) ولا يخص ذلك العرب فقط بل حتّى في اللغات الأخرى «فلم تنتظر المجتمعات البشريّة نهاية القرن التاسع عشر كيما تدرس قضايا الدلالة وتوليها اهتمامها؛ ومن ثمّ لتوظّفها في إطار الفاعليّة المميّزة لعلم اللغة، ولكن العلماء في ميدان النحو وسواء جوانب الدرس اللغويّ أعطوا نتاجا ساعد على معالجة مشكلات دلاليّة منذ الآماد المبكّرة، سواء في المعجمات الّتي بزغت مع الحضارات العربيّة القديمة في سورية وبلاد الرّافدين ومصر والهند وبلاد الإغريق، أو في أعمال اللغويّين والنّحويين ثمّ الفلاسفة وأصحاب الفكر» (الداية، 1996).

يرى محمد المبارك أنّ الدلالة هي إثارة اللفظ للمعنى الذهني أي لمدلوله، الأمر المتواجد في كلّ لغة إثارة متبادلة وتداع مستمر (المبارك، د سنة) . وعليه فالباحث لا يرى أنّ الدّلالة مرادفة للمعنى فقط بل إضافة إلى ذلك هي «العلم الباحث في صلات الألفاظ بعضها ببعض» (المبارك، د سنة).

بالنسبة للنظم الآلية « فتعدّد المعنى للكلمة الواحدة، وحساسيّة السياق في تحديد دلالة الكلمة، واختلاف الدلالة باختلاف الثقافات[...]، كلّ ذلك يجعل المعالجة الآلية للدلالة تنطوي على مفارقات يصعب بسببها تمثيل هذا المستوى أو توصيفه حاسوبيّا، وبسبب من هذا تجاوزت أوّل دراسة صادرة عن اللسانيات الحاسوبية العربية الحديث عن المعالجة الآلية لعنصر الدلالة في العربية » (العارف، 2007)، ولا يعني هذا أنّ جانب الدلالة قد أُغفلَ تماما لكن حضوره كان في عرض المعالجة الآلية للمستويات الأخرى، وقضايا أخرى وثيقة الصّلة بالدلالة كالتّرجمة الآليّة. (العارف، 2007)

أدرج نبيل علي عدّة أسباب للقصور في مناقشة الجوانب الدلاليّة للغة العربيّة أهمّها: (علي، 1988)

التركيز على الهياكل الأساسيّة (التحتيّة) اللازمة لتهيئة اللغة العربيّة للتّعامل مع نظم المعلوماتيّات، فقصرت جلّ الأبحاث عملها على المعالجات اللغويّة الأساسيّة، والّتي تصبّ جميعها في نظم المعالجة الآليّة لعنصر الدلالة.

معالجة الدلالة آليّا بحاجة إلى خلفيّة نظريّة من المنطق والرّياضيّات وأساليب الذكاء الاصطناعي تختلف بصورة كبيرة عن الخلفيّة اللازمة للمعالجات اللغويّة الأدنى.

يعدّ عنصر الدلالة من أقلّ الفروع اللغويّة فيما يخص التباين اللغوي، وتعتمد هذه الأخيرة على التمثيل المنطقي للتّعبيرات اللغويّة، والّذي يتلاشى على مستواه الاختلاف اللغوي بقدر كبير.

المعالجة الدلاليّة آليّا تحتاج إلى الكثير من البحوث النظريّة وأساليب الذكاء الاصطناعي المتطوّرة لمحاصرة زخم المشاكل الّتي تنطوي عليها هذه المعالجة.

القصور الكبير في الدراسات الدلاليّة على مستوى اللغة العربيّة والّذي لا يوفّر الحد الأدنى لتأسيس تصوّرات محدّدة لمعالجة الدلالة العربيّة آليّا.

معرفة أسباب قصور المعالجة الدلالية الآلية للغة العربيّة هو بحدّ ذاته عمل مهم لأنّه يؤدّي بالمهتمّين إلى معرفة الإشكاليّة الحقيقيّة والّتي تعدّ حلا مؤقّتا وممهّدا جيّدا لتفاديه، وجعل هذه الأسباب الّتي كانت نقاط ضعف إلى نقاط قوّة وتطوّر. قد تكون بعض هذه الأسباب صحيح إلى حدّ كبير غير أنّ ما أورده نبيل علي في أنّ ما جاء من أبحاث حول المعالجة الآليّة للغة العربيّة ركّز على الهياكل الأساسيّة فإنّنا نرى أنّ أهم مرتكز كان على الفنيّين التركيز عليه هو المعالجة الدلاليّة الآلية

باعتبارها الأساس الأوّل الواجب البحث فيه لأنّ الباحث ذاته نبّه إلى أنّ المعالجة الدلاليّة آليّا تعمل على محاصرة زخم المشاكل الّتي تنطوي عليها هذه المعالجة، وإلى أنّ التّوصيف الّذي أورده باعتبار موقف المتكلم؛ انطلاقا« من غرض الحديث ومقامه، إلى المضمون (أو المحمول) الدلالي، ومنه إلى اختيار الأساليب والتراكيب النحوية الّتي تنقل المعنى المطلوب، ثمّ اختيار وتكوين الكلمات وإدخال التعديلات التصريفية عليها، ثمّ تطبيق القواعد الفونولوجية لإحداث التنويعات الصوتية والّتي تولّد سلسلة الفونيمات الّتي تحوّلها منظومة «الفونيتيك» في النهاية إلى منطوقات» (علي، 1988) أمّا إذا أوردناه باعتبار المستمع فعمل المنظومة اللغوية يكون بشكل عكسي، عن طريق استخلاص المعاني من الإشارة الصوتية للمنطوق اللغوي من خلال تحديد المعاني الوظيفية للزوائد والحركات الإعرابية، وإمداد الدلالات المعجمية، واستخدام جميع القرائن المعجمية والوظيفية والصوتية والسياقية والمقامية لاستخلاص المعنى المطلوب، وهو ما يؤكّد من جانب آخر أنّ دائرة المعنى هي الحجر الأساس سواء للمتكلم أو المستمع. وأهمّ ما يحلم به علماء التكنولوجيا والبرمجيّات هو كيفيّة التّحدث بالفصحى والعاميّة إلى الحاسب الآلي ويستطيع هو الآخر فهمهم دلاليّا.

وصف أغلب التقنيّين أنّ المعالجة الآليّة للدلالة بالشيّقة لأسباب كثيرة سنذكرها، غير أنّنا نرى أنّ هذه الأسباب لا تعطي عنصر الدلالة وصف المتعة بل تتجاوزها إلى الحتميّة والإلزام باعتبارها هي الانطلاقة الأولى الواجب العمل بها حتّى يتحقّق ما يهدف إليه علماء التكنولوجيا، وتتمثّل هذه الأسباب فيمايلي: (علي، 1988)

- البنية العميقة للغة العربيّة أكثر ملاءمة للتّمثيل الدلالي، ويسهّل من عمل آلة الاستدلال المنطقي.
- شدّة تماسك عناصر منظومة اللغة العربيّة ممّا يزيد من فاعليّة تضافر القرائن اللغويّة المختلفة، ويعدّ هذا التّضافر من العوامل الهامّة في معالجة الدلالة آليّا.
- الصلة الوثيقة بين مباني اللغة العربيّة ومعانيها، والدلالة في جوهرها هي عمليّة ربط الرّمز بمدلوله.
- ينتج عن إسقاط علامات التشكيل حالات لبس يصعب وجود مثيل لها في لغات أخرى، ممّا يجعل نظم الفهم الأوتوماتي للعربيّة غير المشكولة بمثابة حالة تصميم قصوى يمكن أن تندرج تحتها حالات لغويّة أقلّ صعوبة.

تعدُّ مسألة فكّ الالتباس الدلالي للكلمات إحدى المسائل العتيدة والعنيدة الّتي لم يتمّ حلّها بعد بشكل مرض سواء في اللغة العربيّة أم في سواها، وهو ما أدّى إلى اتّساع هامش الخطأ الناتج عن استخدام أساليب لفكّ الالتباس الدلالي للكلمات مقارنة بالأساليب نفسها لفكّ الالتباس الصرفي باستخدام تطبيق أساليب التعلّم الحاسوبي، ولم يقلّ في أمثل الأحوال مع اللغة الإنجليزيّة عن خمسة وعشرين في المئة وقد يرتفع إلى ما فوق الثلاثين في المئة مع اللغة العربيّة، ويمكن تفسير ذلك إلى أنّ السياق الاحتماليّ الدّلالي في النص أكثر اتّساعا في المتوسّط من نظيره الصرفيّ، انطلاقا من أنّ دلالة الكلمة لا ترتبط بما يجاورها من كلمات فقط بل بكامل الفقرة وقد يجاوزها إلى خارج النص، ممّا يتطلّب قدرات أكثر دقّة في المعالجة الحاسوبيّة. (رشوان و آخرون، 2019) وهذا ما يعرقل أداء التطبيقات المرتكزة على التحليل الّدلالي المعجميّ، إنّ هذا النّظر يبيّن لنا مدى صعوبة المهمّة الّتي على عاتق أصحاب البرمجيات واللغويّات، وأنّ العمل سوف يكون شاقا ومتعبا، ويتطلّب دقّة كبيرة.

2–الأنطولوجيا ودلالة اللغة

يرى المنشغلون بميدان العلوم التقنية أنّ أنضج المفاهيم النظامية في التعبير عن التساؤلات الوجودية هو ما عُرف بمصطلح «الأنطولوجيا»(Ontology) ؛ وهو مصطلح فلسفيّ مشتق من لفظة «أُنطو» اليونانية القديمة، ويُعنى بدراسة طبيعة الوجود، وتصنيف الموجودات والعلاقات بينها، وكل ما يخص هذه المباحث عن طريق التمييز بينها، وبغرض الوصول إلى ما قام به الباحثون في مجال اللسانيات الحاسوبية وعلاقته بالأنطولوجيا، فقد استلهم هؤلاء مفهوم المصطلح وتوسّلوا به كإطار لجمع المعرفة والاستفادة منه حاسوبيا، مع الاعتماد من جهة أخرى على ما وصلت إليه «الخوارزمات الحاسوبية» (Algorithms) الحاذقة، فجمعت بذلك المعالجة الحاسوبية مع الأنطولوجيا لتشكّل القدمين السليمتين للذّكاء الحاسوبي، والّذي بإمكانه تجاوز العمليات الصعبة الّتي كانت لا تتخلّى عن الذكاء البشري، ولقد بدأ الاهتمام في مجال معالجة اللغات الحيّة بالأنطولوجيا، ولأنّ العملية ليست بالسّهلة واليسيرة، فإنّه لم يتم الوصول إلى بناء مثل هذه الأنطولوجيا الشاملة لأيّة لغة حيّة، غير أنّ العاملين على حوسبة اللغة بالاستعاضة عن الأنطولوجيا الكاملة ببناء أطر معرفية دلالية للغة حيث تتوسّل بروح الأنطولوجيا ومركّباتها للتمثيل الدلالي كأطر جزئية أنطولوجيّة. (رشوان و آخرون، 2019)

وعلى الرغم من أنّ مصطلح الأنطولوجيا فلسفيّ المولد والنشأة والدلالة اللغويّة، فهذه المعرفة من أحدث التطبيقات الحاسوبية لمعالجة الدلالة للغات الطبيعيّة، حيث تتكوّن من عناصر أهمّها: (سليمان و آخرون، 2017)

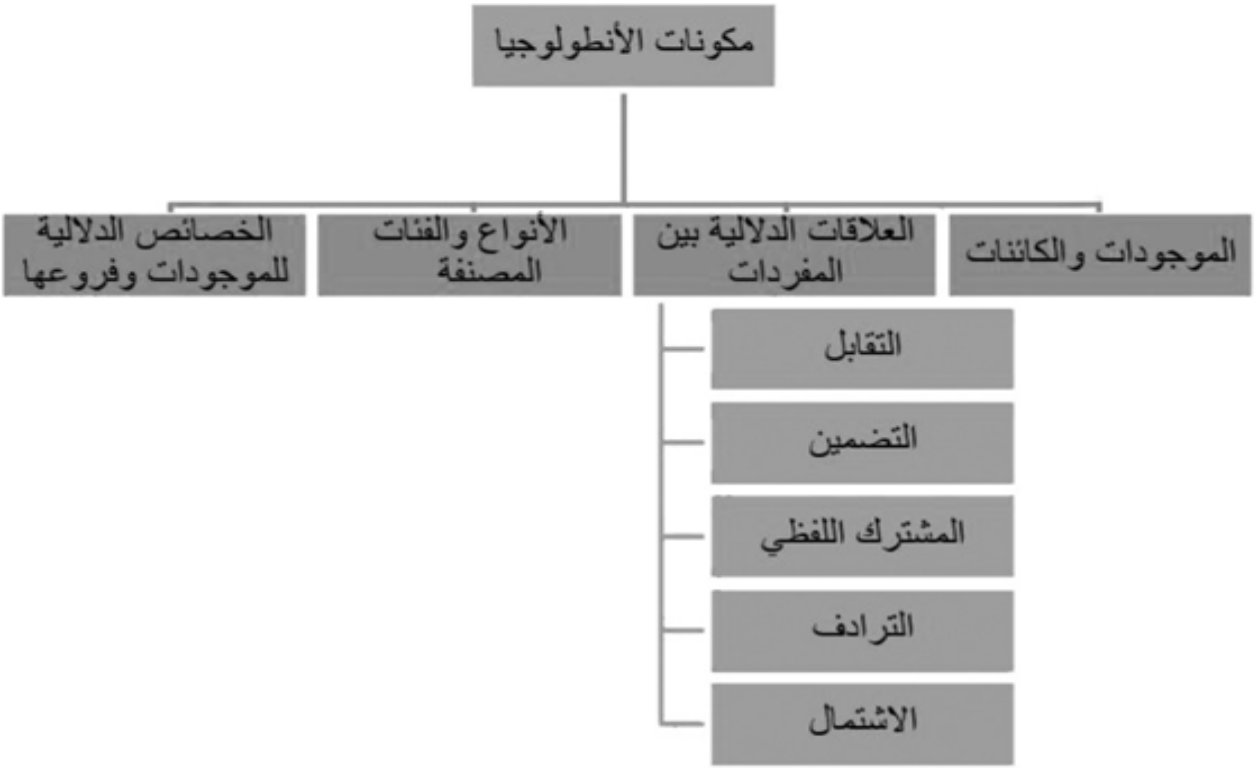
- الموجودات والكائنات.
- الأنواع والفئات المصنفة والطبقات تحتها.
- الوظائف وتشمل (الحدث والمنفذ والأداة والمكان).
- السمات الدلالية للموجودات وفروعها.
- العلاقات الدلالية بين الموجودات وفروعها.

تعدّ هذه العناصر عناصر ثابتة، وقد أضاف بعضهم عناصر أخرى نعتبرها غير ثابتة تتمثّل فيمايلي: (رشوان و آخرون، 2019)

- الأحداث؛ حيث يشير الحدث إلى تغيّر في خصائص الأفراد/ أو الموجودات و/الفئات، و/ العلاقات بين أيٍّ منها.
- الوظائف؛ وهي الّتي تسمح بتكوين بِنى مركّبة من الأفراد يمكن التّعامل معها لاحقا كوحدات أنطولوجيّة.
- القيود؛ وتوضع على ما يمكن قبوله كمدخلات من صيغ تقريريّة لحالة الأنطولوجيا أو جزء منها.
- القواعد (الشرطيّة غالبا- في شكل مقدّمات ونتائج) الّتي تصف الاستدلال المنطقي الّذي يمكن استشفافه من مجموعة الصيغ التقريريّة المقبولة.
- المسلّمات.

تنهض بهذا التصنيف ثلاث نظريات دلالية حديثة هي: نظرية الحقل الدلالي والّتي تعد المفاتيح الرئيسة الّتي يستدل بها على الكلمات، ونظرية التحليل الدلالي والّتي تعمل على تحليل المعنى في عدّة مستويات مع تبيان العلاقات الدلالية بينها، أمّا نظرية العلاقات الدلالية فهي ترتبط بعمل النظرية الأولى، فمعنى الكلمة يتحدّد من خلال علاقته بالمفردات المتواجدة معه في ذات الحقل الدلالي (سليمان و آخرون، 2017)، وخدمة النظريات اللغوية الحديثة للأنطولوجيا يظهر مدى تجانس عملهما، والّذي من شأنه أن يصل إلى نتائج تخدم اللغة العربية.

الشكل المبيّن أدناه يمثّل الموكوّنات الرّئيسيّة والثابتة في بناء الأنطولوجيا حسب ما جاء عند غالبيّة باحثي البرامج: (سليمان و آخرون، 2017)



شكل1: مكونات الأنطولوجيا

المصدر: محمّد رشوان وآخرون، مقدّمة في حوسبة اللغة العربيّة.

يرجع الاهتمام من قبل التقنيين بالأنطولوجيا على أساس أنّها تستطيع تشكيل الإطار الرّابط بين الطبقات الأوّليّة لمعالجة اللغة الحيّة؛ كالتحليل الصوتي والصّرفي، وبين الطبقات العليا لمعالجة اللغة؛ كـ: الترابط الخطابي (Discourse Integration) ومعرفة:

المغزى السياقي للكلام (Pragmatics) وهي الّتي تستلزم معرفة بواقع العالم وتفاعلاته. (رشوان و آخرون، 2019)

تهدف العناصر المكوّنة للأنطولوجيا إلى بناء مرجع رقميّ موحّد للمصطلحات والمفاهيم والعلاقات في مجال معيّن، بهدف استخدامها في تبادل ونشر المعلومات بين المختصين أو في البرامج الحاسوبيّة لإزالة اللّبس في المفاهيم. ومن خلال جهود الباحثين في مجال الذكاء الاصطناعي فقد تمّ تفعيل المنطق الرّياضي لاستنتاج حقائق ضمنيّة ومعلومات جديدة لبناء قواعد معرفيّة حاسوبيّة. وقد تضافرت الجهود لإرساء ما يعرف بـ: لغة الأنطولوجيا الشبكيّة (Ontology Web Language)، وقد شاعت هذه اللغة مع شيوع مشاريع الويب الدلالي (Semantic Web) وتطبيقاتها. ويجدر التنبيه إلى أنّ علوما مختلفة وظّفت الأنطولوجيا بمفهومها الحاسوبي لتمثيل مفاهيمها الخاصة كالطبّ لتوصيف الأمراض والأدوية، والأحياء لتوصيف الجينات، والكيمياء لتوصيف المركّبات، وعلم المكتبات لتوصيف المؤلّفات، واللغويّات لتوصيف الكلمات، وكذلك في كلّ علم أو تطبيق يحتاج إلى توصيف المفاهيم والعلاقات بشكل مقنّن ودقيق. (سليمان وآخرون، 2017)

بدأ علم الأنطولوجيا يأخذ مكانته حديثا لارتباطه بهندسة المعرفة والتدبير الدلالي الحاسوبي، ولن يتأتّى لهذا العلم جني ثماره إلا بوصله بالدّراسات اللسانيّة وما وصلت إليه في مفهوم الدلالة وفكّ غموض المعنى. وعليه فالدلالة هي القلب النابض للأنطولوجيا وأيُّ فتوحات جديدة هو فتح لها، ومن شأن ذلك أن يمنح اللسانيّات الحاسوبيّة فرصا أوسع لخدمة الشبكة الدلاليّة والمحتوى اللغوي حاسوبيّا.

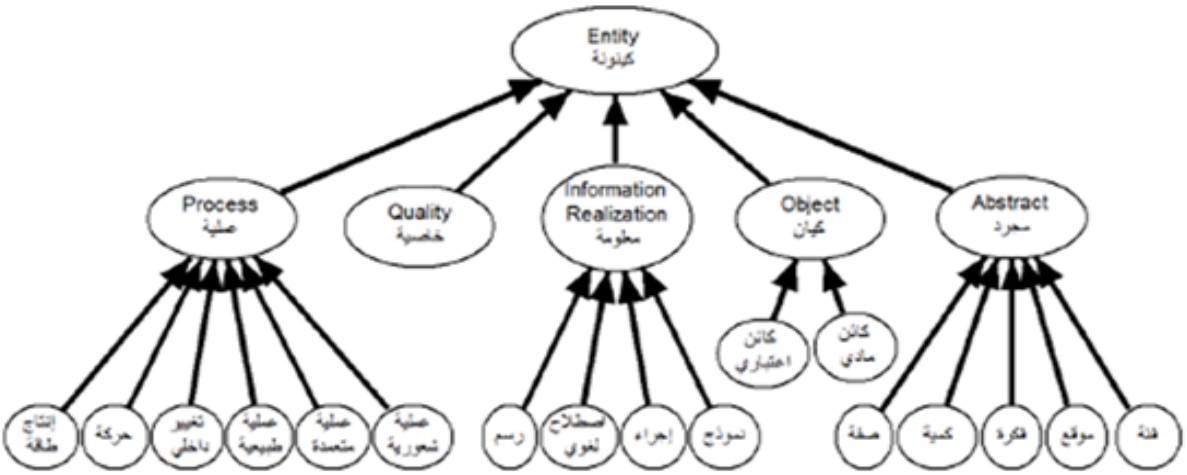
3–الأنطولوجيا وقاعدة البيانات المعجميّة

سعيا وراء تحقيق حوسبة اللغة العربية فإنّ أصحاب التقنيّة يرون أنّ الأنطولوجيا هي الحجر الأساس للتّبادل السليم والفعّال للبيانات، حيث تحتوي على تعريف دقيق للمعنى الدلالي للبيانات المراد تبادلها. حيث تُكتب هذه التعريفات بلغة المنطق كي يستطيع أيّ نظام فهمها وحسابها، بل والاستنتاج منها. وقد ظهر في السنوات العشر الأخيرة الكثير من التطبيقات الّتي تعتبر فيها الأنطولوجيا بالغة الأهميّة مثل الحكومات الإلكترونية، التجارة الإلكترونيّة، محرّكات البحث، المكتبات الإلكترونية، وغيرها من التطبيقات. (جرار، 2011)

أنطولوجيا اللغة الإنجليزيّة (WordNet) كانت بمثابة الخطوة الأولى لإنشاء الأنطولوجيّات اللغويّة في مجال الأبحاث، وحذا حذوها الكثير من اللغات الأخرى، غير أنّ هذه الأنطولوجيّات لم تكن مشجّرات مفاهيميّة منفصلة بل تمّ ربط المفاهيم المتطابقة عبر اللغات، بمعنى المفهوم نفسه يحيلنا إلى رقم واحد بغضّ النظر عن اللغة الأصليّة له. أمّا بالنسبة للغة العربيّة فقد كانت البداية الأولى للمشروع الّذي قامت به وكالة الاستخبارات الأمريكيّة غير أنّه لم يتمّ تصنيف معاني الكلمات بالكامل، وقد انتهج هؤلاء أسلوب ترجمة أنطولوجيا اللغة الإنجليزيّة بما يقابلها في العربيّة، ولم تكن هذه المنهجيّة سبيلا سليما ونموذجا فعّالا لأنّ الأنساق الفكريّة واللغويّة للمفاهيم اللغويّة لا يمكن أن تتطابق في بنيتها الإنتاجيّة ودلالاتها بشكل حرفيّ. (جرار، 2011) رغم وجود تطابقات كثيرة بين مختلف اللغات.

بناء أنطولوجيا للّغة العربية يحتاج إلى العديد من السنوات والجهود البحثيّة، كما تعدّ الانطلاقة البحثيّة الجادّة لمشاريع مستقبليّة طويلة الأمد، وكما يؤكّد أصحابها فهي تؤسّس لطريقة جديدة في تعريف معاني ودلالات الكلمات، وتهدف إلى إنتاج قاموس دلالي آلي تصويري يصنّف هذه المعاني ويشجّرها، فتكون هذه المعاني والعلاقات مؤصّلة فيما بينها وممثّلة بلغة المنطق الشكلي فلسفيّا ولغويّا (جرار، نحو تأصيل منهجي لبناء أنطولوجيا اللغة العربية، 2011) ويتلخّص ما تمّ إنجازه في هذه المنهجية فيما يلي: (جرار، 2011)

- بناء مستويات عليا لأنطولوجيا اللغة العربية (Level Top Concepts)، والّتي تشكل نواة الأنطولوجيا العربيّة، والّتي يتمّ تأصيلها تأصيلا فلسفيّا داخليا ومنذ البداية بالاعتماد على الأنطولوجيات العليا العامّة (Upper Level Ontolgies) خلافا لما يحدث في الأنطولوجيا الإنجليزيّة (WordNet) والّتي قد تمّ ربطها ربطا خارجيا بعد استكمالها. وهو عمل من شأنه تغييب الاتساق بشكل كليّ.
- جمع واستنباط تعريفات ومعاني من القواميس العربيّة المتاحة وإعادة صياغتها وهندستها كتعريفات دلاليّة تخضع لضوابط تركز على صفات جوهريّة مميّزة للمفهوم دون غيره، وقد تمّ تطبيق ذلك على ما يقارب من الثلاثين ألف مفهوم.
- تطوير برنامج حاسوبي مبني على خوارزميّة ذكيّة عملها الأساسي الربط بين مفاهيم الأنطولوجيا العربيّة
- ومقابلاتها في الأنطولوجيا الإنجليزيّة، (WordNet)، الأمر الّذي يتيح جلب علاقاتها الدلاليّة إلى العربية.



شكل 2: جزء من المستويات العليا للأنطولوجيا

المصدر:

Mustafa Jarrar: Building a Formal Arabic Ontology (Invited Paper). Proceedings of the Experts Meeting on Arabic Ontologies and Semantic Networks. ALECSO, Arab League. Tunisia. July, 2011/http://www.jarrar.info/

يصف مصطفى جرار هذه العمليّة بالشاقّة إذ تستوجب التّحقق من أنّ المعاني المعرّفة وجميع التطبيقات صحيحة بشكل عام، وأهم ما تتضمّنه الأنطولوجيا للتّدليل والإشارة إلى معنى ما هو التّصنيف. غير أنّ هذه الهندسة تحمل بين ثناياها العديد من الإشكالات المنطقيّة والمنهجيّة. وتجنّبا للمشاكل الّتي يواجهها استخدام التصنيفات للتّدليل على المعاني في الأنطولوجيا، ورغم شيوع هذه الأخيرة اقترح فريق البحث منهجيّة جديدة للتّدليل على المعنى، مفادها أنّها تعتمد على صفات الأشياء وليس على أجناسها، ويصفها أصحابها بالسّهلة ذات النتائج السليمة والاتّساق دلالي (Ontology consistency). فبدل من تصنيف الأشياء إلى صنف عام وآخر خاص برزت فكرة استخدام الصفات المميّزة للأشياء للتّدليل على معانيها.

يستفيد من الأنطولوجيا أطراف عديدة فعلى سبيل المثال مجال الحكومة الإلكترونيّة الفلسطينيّة، يجب أن تتضمّن الأنطولوجيا على جميع المفاهيم المستخدمة في قواعد جميع مؤسسات الحكومة الفلسطينيّة، مفاهيم مثل: شخص، مواطن، شركة ...، وغيرها، ويتمّ إعادة تعريف هذه المفاهيم بدقّة ضمن نسق شجريّ تصنيفي، وتعدّ بذلك هذه الأنطولوجيا المرجع الدلالي للرّبط المفاهيمي بين أنظمة المعلومات في هذه المؤسسات، يسهّل التبادل البيني الآلي ذو قاعدة موحّدة للفهم (Meaningful Interoperation). وقد استفاد معهد الحقوق لجامعة بيرزيت لإغناء برنامج المقتفي وهو محرّك بحث متخصّص في مجال القانون الفلسطيني. (جرار، 2011) ويجدر الإشارة إلى أنّ التصنيف في الأنطولوجيا لا يبنى على أساس الترادف اللغوي كما في القواميس، أو الخاص والعام كما في المكانز، بل يتمّ بناؤه على صفات الجنس ونوعه باستخدام المنطق الشكلي؛ بمعنى تصنيف المعاني وتحديد وتمييز تعدد معاني كلّ كلمة، ولذلك فعلاقات الأنطولوجيا هي علاقات دلاليّة وليست لغويّة والّتي نقصد بها: المشتق، واسم الفاعل، والمصدر، وصيغة المبالغة، وغيرها من العلاقات اللغويّة بين الكلمات. (جرار، 2011)

إنّنا بحاجة لبناء إطار دلالي تأسيسي حاسوبي للغة العربيّة يمكّنها من اللحاق بالرّكب اللغوي لباقي لغات العالم، وفق منهجيّة بحث وصفيّة تحليليّة ضمن خمس مكوّنات، تشكل خطوات البحث والعناصر المنهجيّة والتقييميّة الخاصّة بها، وهي كالآتي في مايلي: (جرار، 2011)

3-1-الخطوة الأولى: جمع واستنباط تعريفات ومعاني من القواميس العربيّة المتاحة المتخصّصة والعامة

بغرض جمع أكبر عدد ممكن من المعاني المتعدّدة للكلمة الواحدة في عدّة مجالات، غير أنّ هذه العمليّة لا تتمّ مع جميع أصناف المصادر اللغوية (معاجم، مكانز، مسارد، أنطولوجيات)، فتُبعد تلك الّتي تركّز على تصريف الكلمات، وتُخلط بين التصريف اللغوي والتعدد الدلالي، وأمثلتها إيمائيّة بمعان غير واضحة، ولذلك فلابدّ من توفّر مجموعة من المعايير في اختيار صنف المصادر اللغويّة للمعاني

الدلاليّة، وهي كالآتي:

- ضرورة بناء المعجم على المعاني الدلاليّة وتعدّدها بالنسبة لكلّ مصطلح.
- وضوح المعنى ودقّته.
- جودة التعريف وطريقة تركيبه، بحيث يكون مكتوب بطريقة واضحة وكلمات صحيحة.

وفرة المصادر اللغويّة تمكّن من بناء تطبيقات حاسوبية، وقد نبّه الباحثون في البرمجة إلى شحّ المصادر اللغويّة المحوسبة إلى درجة أنّه لا يوجد قائمة بجميع المدخلات العربيّة، إضافة إلى غياب مصادر لغويّة حديثة فمثلا لا يوجد مكانز مترادفات لأغلب الكلمات بتصميم جيّد، وإن وجدت فهي قصيرة لا تغطي الحجم المناسب. ولذلك فهذه الأسباب تشكّل ضعفا كبيرا في دعم تطبيقات للغة العربيّة.

أخذت التجربة الفلسطينيّة على عاتقها عمليّة جمع ورقمنة وتنقيح ودمج وتوحيد المصادر اللغوية، وتعمل على إتاحتها بطريقتين اثنتين: للجمهور العربي والباحثين ومتعلّمي اللغة العربية خاصة ومطوّري التطبيقات عبر واجهات برمجيّة.

والشكل المدوّن أسفله يظهر أصناف المصادر اللغويّة في جامعة بيرزيت: (.«Extracting Synonyms from Bilingual Dictionaries»,» 2021)

شكل 3: شبكة بيانات لغويّة ضخمة تجمع المستويات التصريفية والاشتقاقيّة والدلاليّة للفصحى والعامية

المصدر:

Mustafa Jarrar.«Extracting Synonyms from Bilingual Dictionaries» . Research Talk. The 11th Global Wordnet Conference. SouthAfrica212021/1/.http://www.jarrar.info/

مدوّنة للعاميّة قاعدة بيانات معجميّة الأنطولوجيا العربيّة

مجموعة ضخمة من النصوص العاميّة، أضخم قاعدة بيانات معجميّة في تاريخ تصنيف مفاهيم(معاني) الكلمات تمّ تصريف ووسم كل كلمة فيها بحوالي العربيّة (تصريفات، اشتقاقات، دلالة) العربية (عمل فلسفي حاسوبي) 16سمة

يبيّن الشكل أعلاه أنواع المصادر اللغويّة في جامعة بيرزيت، مدوّنة للعاميّة بلهجات مختلفة، وقاعدة بيانات معجميّة عملاقة تضمّنت أكثر من 150 معجما من مستويات مختلفة كالترجمة والتصريفات والاشتقاقات والتعريفات والمترادفات ... إلى غيرها من المستويات، والأنطولوجيا العربية وهي المصدر النّواة، وعند ربط هذه المصادر بعضها ببعض يتشكل لنا شبكة بيانات لغويّة ضخمة لجميع المستويات التصريفيّة والاشتقاقيّة والدلاليّة الفصحى والعاميّة.

وتتمّ هذه العمليّة يدويّا أو باستخدام الـ(OCR). الماسح الضوئي لقراءة النصوص آليّا، وإتباعها بالمراجعة اليدويّة للتّأكّد من صحّتها، وقد قام أعضاء العمل عام 2011م باستنباط ثلاثين ألف تعريف، ويُتوقّع الوصول إلى ما يُقارب السبعين ألف في العام الموالي من معاجم كثيرة ومتنوّعة.

3-2-الخطوة الثانية: إعادة صياغة وهندسة التعريفات(glosses) المستنبطة في الخطوة الأولى

بطريقة يدويّة من خلال الضوابط الأنطولوجيّة والّتي تركّز على الصّفات الجوهرية المميّزة المتمثّلة فيمايلي:

- ضرورة بدء التعريف بالجنس الأعلى للمفهوم المراد تعريفه؛ كبداية تعريف «المصفوفة» بـ «ترتيب...».
- ذكر جميع الصفات الجوهريّة المميّزة للمفهوم (تستبعد الصفات العرضية أو الاشتقاقات اللغويّة).
- كتابة الصفات الجوهريّة بطريقة تصوّريّة تقود لاستنباط المفهوم.
- الإشارة إلى تصحيح ما شاع خطأه، وتخطيء ما شاع صحّته.

- تصنيف التعريف المدرج ضمن الشجرة المفاهيميّة.

- الوضوح والاختصار شرط ضروري في التعريفات.

من شأن هذه الشروط أن تضفي صفة الدقّة والشّموليّة، والحصول على تعريفات دلاليّة ذات منطق صارم بمنهجيّة بناء هندسة الأنطولوجيات (التصنيف بالصّفات الجوهريّة).

3-3-الخطوة الثالثة: ربط التعريفات المنجزة في الخطوة السابقة بما يُقابلها في أنطولوجيا اللغة الإنجليزية إن وُجد

ولم يتحقّق ذلك إلا عن طريق برنامج ذكي بشكل آلي بدقة 90 %، ويجدر الإشارة إلى أنّ هذا البرنامج يعمل على البحث عن تعريفات مكرّرة أو متداخلة، وتتلخّص خوارزميّة هذا البرنامج كمايلي:

- ترجمة آليّة للتّعريف العربي باستخدام Google Translate، أو محرّكات ترجمة أخرى متاحة.
- إضافة جميع الاشتقاقات والمترادفات لكل كلمة واردة في ترجمة التعريف، وجميع الكلمات ذات العلاقة الدلاليّة أو اللغويةّ الاشتقاقيّة المرتبطة، وقد يصل عددها إلى المئات وأحيانا الآلاف.
- مقارنة جميع الكلمات بكلمات موجودة في قائمة تعريفات الأنطولوجيا الإنجليزيّة، وتُعطى علامة لكل تطابق يحدث في هذه الكلمات.
- حساب مدى تقارب التعريف الأصلي المُدخل بكل تعريف إنجليزي، ويُرشّح التقارب الأعلى نسبة ليكون هو المقابل الانجليزي للمفهوم العربي، والعمليّة عكسيّة، وقد وصلت دقّة الربط إلى 90 %.

3-4-الخطوة الرابعة: بناء العلاقات الدلاليّة بين التعريفات ضمن الشجرة المفاهيميّة

أي أنّ نجاح هذه الخطوة له علاقة مباشرة بنجاح سابقتها، فالنّجاح في تطبيق الضوابط الأنطولوجية وهي بدء التعريف بالجنس الأعلى منه يقود إلى تعريف علاقة جنس من جنس لـ(SuperType/SubType) بين مفهومين أو أكثر من جهة، ومن جهة أخرى النّجاح في الربط الدلالي بين المفهوم العربي ومقابله الإنجليزي يؤدّي إلى استنباط معظم العلاقات الدلاليّة من الأنطولوجيا الإنجليزية.

3-5-الخطوة الخامسة: ربط المفاهيم والتعريفات المنتجة في الخطوة الثانية والعلاقات المنتجة في الخطوة الرابعة بالمفاهيم العليا (Top-Level Concepts) للّغة العربيّة

فبناء على ما تمّ بناؤه من شجرة مصغّرة تدعى الشجرة العليا والّتي تتكوّن من عشر مستويات من المفاهيم العليا الأم لجميع مفاهيم اللغة العربية، حيث تُستخدم وفق الأهداف الآتية:

- - تُستخدم الشجرة العليا النواة الأم لربط جميع المفاهيم بها؛ فكل مفهوم منتج يتمّ ربطه بإحدى المفاهيم في المستوى الأخير من الشجرة العليا والّذي يعلو جميع المستويات المنتجة في الخطوة الرابعة.
- - تُستخدم الشجرة العليا النواة للتّحقّق من صحّة التعريفات والعلاقات المنتجة سابقا لضمان الجودة العالية في عمليّة التشجير (تصنيف المعاني) كأن لا يكون للمعنى أكثر من جنس واحد يعلوه ما أمكن ذلك، فتعدد

الأجناس ناجم عن عدم الفهم والدقّة، ممّا يضمن ويوفّر التحكم في صحّة العلاقات الدلاليّة.

شكل4: خطوات البحث والعناصر المنهجيّة في بناء الأنطولوجيا العربية.

المصدر: المؤلف

4-حوسبة المعاجم العربيّة

عرض فريق البحث لجامعة بيرزيت تجربةً لبناء أضخم قاعدة بيانات لغويّة للغة العربيّة، إضافة إلى أوّل محرّك بحث للمعاجم العربيّة عالميّا يتضمّن 150 معجما تمّت حوسبتها، وقد عمل الفريق على أن يكون محرّك البحث هذا متاحا للعامّة ولجميع الطلبة والباحثين والمترجمين ومتعلّمي اللغة العربيّة وغيرهم ممّن هم بحاجة لمثل هذه الأبحاث. يتيح محرّك البحث المعجمي للمستخدم البحث عن كلمة ضمن 150 معجما عربيّا ومتعدد اللغات مع استرجاع تعريفاتها ومرادفاتها وترجماتها المتخصّصة، ولا يستخدم محرّك البحث المعجمي الترجمة الآليّة المتاحة مثل جوجل للتّرجمة بل هي ترجمة تعتمد على المعاجم المدقّقة والمنقّحة. ويجدر الإشارة إلى أنّ هذا المحرّك هو الأوّل عالميّا فلا يوجد مثله في اللغات الأخرى حتّى الإنجليزيّة. وقد تمّ حوسبة المعاجم بشكل يدوي خلال تسع سنوات في قاعدة بيانات واحدة في ظلّ غياب نسخ إلكترونيّة لغالبيّة المعاجم حتّى في حوزة مؤلّفيها. اختلفت المعاجم اللغويّة بين القديم والحديث، والمسارد، والمكانز، ومعاجم ثنائيّة وثلاثيّة اللغة، وقواعد بيانات تصريفيّة واشتقاقيّة ترتبط ارتباطا مباشرا بالأنطولوجيا العربيّة.. (جرار، حوسبة المعاجم العربيّة -دراسة حالة-، 2017) (جرار، محرّك بحث للمعجم العربي، 2018) **وما** يزال فريق البحث يعمل على ذلك للوصول إلى شبكة لغويّة محوسبة للغة العربيّة تسمّى: Big linguistic Data draph وربطها باللغات الأخرى.

من خلال موقع جامعة بيزيت الإلكتروني (https://ontology.birzeit.edu) يمكن لأيّ مستخدم إيجاد المصطلح العربي الّذي يوافق متطلّباته في شتّى المجالات العلميّة والهندسيّة والتّجاريّة والأدبيّة وغيرها. تستفيد فئات أخرى من محرّك البحث المعجمي إضافة للباحثين واللغويّين والمترجمين والطلبة والجمهور العربي أيضا الشركات الّتي تحتاج لبيانات لغويّة من أجل تطوير تطبيقات حاسوبيّة كالترجمة الآليّة، والبحث الدلالي، والتحليل اللغوي، والمدققات الإملائيّة، واستخراج البيانات وغيرها. (جرار، حوسبة المعاجم العربيّة -دراسة حالة-، 2017) والشكل أدناه هو الصفحة الّتي يمكن من خلالها الدخول إلى محرّك البحث.

شكل 5: صورة الشاشة لصفحة المحرّك المعجمي

المصدر: http://ontology.birzeit.edu/

يقدّم محرّك البحث اسم المعجم ورمز حقوق الملكيّة بجانب البيانات الّتي يتمّ استرجاعها، وعند الضغط على اسم المعجم يظهر اسم المؤلّف والنّاشر، وروابط إلى صفحاتهم الإلكترونيّة، وكذلك صفحة لشراء النسخة الورقيّة للمعجم. ولقد أضافت هذه السمات حظوظا أكثر للفريق من أجل حوسبة اللغة وهندسة المعاجم فقد عقدت اتّفاقا مع منظمة الألسكو للتّعاون في هذا المجال، حيث مُنحوا حوالي خمسين معجما ثلاثي اللغة صادرة عن مركز التعريب بالرّباط. (جرار، 2018)

الشكل أدناه هو مثال عن طريقة استرجاع ترجمات كلمة «attribute» من عدّة معاجم ومن الأنطولوجيا، وللمستخدم الحريّة في استرجاع الترجمات أو مترادفات أو تعريفات حيث تظهر له على الجهة اليمنى نتائج الأنطولوجيا، ونتائج المعجم على الجهة اليسرى.

أثناء البحث في محرّك البحث المعجمي فإنّه يوفّر لنا المعلومات الكاملة للمدخلات المعجميّة مع معاجمها، ونستطيع من خلال النقر على أيّ معجم معرفة المعلومات الكاملة الخاصة

به من فكرة بنائه ومنهجه، ويوفّر إمكانيّة شراء النسخة الإلكترونيّة مع المحافظة على حقوق الملكيّة لأصحابها.

شكل 6: صورة للواجهة الأماميّة لمحرّك البحث

المصدر: http://ontology.birzeit.edu/

شكل 7: صورة لنتائج بحث كلمة «الدلالة» مع إظهار مختلف المعاجم وبياناتها

المصدر: https://ontology.birzeit.edu/term

من أهمّ السّمات الّتي يرتكز عليها هذا العمل هو الرّبط المفاهيمي وهو الرّبط الّذي يحقّق الاتساق بين المفاهيم ولعلّ هذه السّمة هي الّتي جعلت من البحث ناجحا إلى حدّ كبير.

لا يقتصر هذا العمل على اللغة العربية فقط بل أيضا العاميّة الفلسطينيّة وقد أضيفت لهجات عربيّة أخرى كالليبيّة واليمنيّة واللبنانيّة وغيرها، والغاية من ذلك تمكين الحاسوب من فهم النصوص العربيّة سواء كانت فصيحة أو عاميّة. إنّ عمل محرّك بحث معجمي لا يخدم العربيّة فقط بل أيضا المعاجم العربية ليتمّ حوسبتها إلكترونيا وربطها بالأنطولوجيا، فهذا العمل الدقيق قد ألّف بين المعاجم ومحرّكات البحث واللغة العربية والفلسفة التحليليّة.

لقد اتّبعنا في الاشتغال بمسالك الحفاظ على الموروث بمنهجيّة تستوجب منّا النظر مليّا الوقوف أمام هذا الجهد المعرفي الّذي قضى في إنتاجه جهابذة العلم على مرّ القرون، ومن واجب أهل الاختصاص في كلّ مجال علمي إحياء التراث في عقول الأجيال القادمة، ومسالك هذا الإحياء تتوسّل بما وصلت إليه مستجدّات التّطور التكنولوجي لأنّها وببساطة مفروضة علينا، ولابدّ من استغلال هذه المزيّة في صالح العلم لا ضدّه، فمن غير الممكن أن يضرب العلم بعضه بعضا إلا إذا أسأنا استخدامه، ووفقا لهذه الاستراتيجيّة نكون قد طرقنا آفاقا جديدة في إحياء التراث تحمل على استئناف عطاء موروثنا الحضاري وفق سبل تكنولوجيّة حديثة، ومن خلال ذلك ندعو إلى:

- أن تجد مثل هذه الأبحاث الجادّة أو كما تمّ تسميتها من قبل أهلها بالصّناعة اهتماما وعناية لائقة خاصة وأنّها تحقق الأمن اللغوي واقتصاد المعرفة؛

- ضرورة تشجيع الاستثمار بتوفير فرص عمل لجهات عديدة بين اللغويين، والبرمجيين، ومتخصصي اللغة العربيّة أولى من غيرهم في تنمية اقتصاديات دولهم وترقية لغتهم الأم؛

- الاستفادة من محرّكات البحث المعجميّة في هياكل الدولة ومؤسساتها؛

- محاكاة التجربة الفلسطينيّة في جامعات الدول العربيّة فيما يخص اللهجات العربيّة رغبة في تشجيع وبعث التنمية السياحيّة

من خلال توفير تطبيقات ذكيّة تسهّل عمليّة التواصل في تلك البلدان وترويج التراث الثقافي.

المراجع

1. http://www.jarrar.info.
2. .«Extracting Synonyms from Bilingual Dictionaries». (2021). South Africa.
3. عبد الرّحمن الحاج صالح. (2012). بحوث ودراسات في اللسانيّات العربيّة (الإصدار د ط). الرغاية، الجزائر: المؤسّسة الوطنيّة للفنون المطبعيّة.
4. عبد الرحمن بن حسن العارف. (2007). توظيف اللسانيات الحاسوبية في خدمة الدراسات اللغوية العربية- جهود ونتائج-. مجلة اللسانيات ، المجلد 12 (العدد 01).
5. عمر مهديوي. (2012). دور قاعدة المعطيات في بناء المعجم الإلكتروني العربي. مجلّة الصوتيّات (العدد 12).
6. فايز الداية. (1996). علم الدلالة العربي: (دراسة تاريخيّة، تأصيليّة، نقديّة) (الإصدار 2). بيروت، لبنان: دار الفكر المعاصر.
7. محمد المبارك. (د سنة). فقه اللغة وخصائص العربيّة (الإصدار د ط). بيروت، لبنان: دار الفكر.
8. محمد رشوان، وآخرون. (2019). مقدمة في حوسبة اللغة العربية (الإصدار 1). الرياض، السعوديّة: دار وجوه للنشر والتوزيع.
9. مصطفى جرار. (2017). حوسبة المعاجم العربيّة -دراسة حالة-. الإمارات العربية المتحدة: وزارة الشباب والرياضة.
10. مصطفى جرار. (2018). محرّك بحث للمعجم العربي. تم الاسترداد من http://ontology.birzeit.edu
11. مصطفى جرار. (2011). نحو تأصيل منهجي لبناء أنطولوجيا اللغة العربية. تونس: المنظمة العربيّة للتربية والثقافة والعلوم، جامعة الدول العربيّة.
12. نبيل علي. (1988). اللغة العربيّة والحاسوب (دراسة بحثيّة) (الإصدار دط). الكويت: مؤسّسة تعريب.
13. هند بنت سليمان، وآخرون. (2017). علم الدلالة والأنطولوجيا من منظور حوسبة اللغة العربية (الإصدار 1). الرياض، السعودية: دار وجوه للنشر والتوزيع.

Toward an automated architecture for building Arabic ontologies and lexical search engines - Bzeit Palestine University as an example-

Abstract

Among the civilizational challenges and historical responsibilities is what our Arabic language is facing in completing and applying computational linguistics. It is thus preparing for a qualitative societal shift that requires the owners of this language to reveal its characteristics in order to facilitate automated processing, which is what most countries have done with their languages and dialects, making it necessary for us to develop our language through artificial intelligence methods and the automated ability to deal with the language. Since automated processing can only deal with what is clear and specific, this matter constitutes an obstacle and an important milestone in the future of the Arabic language, which made its specificity face an issue at the level of semantics, which is difficult to control, and thus became an issue faced by databases in order to achieve successful digitization by all standards, and therefore we present this research paper to identify this issue through the experience of "Birzeit University" according to a descriptive study. Through this study, in which we relied on the new technology and its employment in the computerization of the Arabic language, we tried to highlight the role of electronic dictionaries in order to model and represent knowledge in a sound systematic way that facilitates computer processing, and in order to achieve this, building Arabic ontologies is the basis for the proper and effective exchange of data, through an accurate definition of the semantic meanings to be exchanged.

Keywords
Ontology
lexical semantics
computational linguistics
computer
electronic dictionaries

Vers une architecture automatisée pour la construction d'ontologies arabes et de moteurs de recherche lexicale - l'exemple de l'Université Bzeit Palestine -

Résumé

Parmi les défis civilisationnels et les responsabilités historiques, notre langue arabe est confrontée à l'achèvement et à l'application de la linguistique computationnelle. Elle se prépare à un changement sociétal qualitatif qui exige que les propriétaires de cette langue révèlent ses caractéristiques afin de faciliter le traitement automatisé, ce que la plupart des pays ont fait avec leurs langues et leurs dialectes, ce qui nous oblige à développer notre langue grâce à des méthodes d'intelligence artificielle et à la capacité de traitement automatisé de la langue. Comme le traitement automatisé ne peut traiter que ce qui est clair et spécifique, cela constitue un obstacle et une étape importante dans l'avenir de la langue arabe, ce qui a fait de sa spécificité un problème au niveau de la sémantique, qui est difficile à contrôler, et est donc devenu un problème auquel sont confrontées les bases de données afin de parvenir à une numérisation réussie selon toutes les normes, nous présentons ce document de recherche pour identifier ce problème à travers l'expérience de l'«Université Birzeit de Palestine» selon une étude descriptive. A travers cette étude, dans laquelle nous nous sommes appuyés sur la nouvelle technologie et son emploi dans l'informatisation de la langue arabe, nous avons essayé de mettre en évidence le rôle des dictionnaires électroniques afin de modéliser et de représenter la connaissance d'une manière systématique solide qui facilite le traitement informatique, et pour y parvenir, la construction d'ontologies arabes est la base pour l'échange approprié et efficace des données, à travers une définition précise des significations sémantiques qui doivent être échangées.

Mots clés
ontologie
sémantique lexicale
linguistique informatique
ordinateur
dictionnaires électroniques

Competing interests

The author(s) declare no competing interests

تضارب المصالح

يعلن المؤلف (المؤلفون) لا تضارب في المصالح

Author copyright and License agreement

Articles published in the Journal of letters and Social Sciences are published under the Creative Commons of the journal›s copyright. All articles are issued under the CC BY NC 4.0 Creative Commons Open Access License).
To see a copy of this license, visit:
http://creativecommons.org/licenses/by/4.0/
This license allows the maximum reuse of open access research materials. Thus, users are free to copy, transmit, distribute and adapt (remix) the contributions published in this journal, even for commercial purposes; Provided that the contributions used are credited to their authors, in accordance with a recognized method of writing references.

© The Author(s) 2023

حقوق المؤلف واذن الترخيص

إن المقالات التي تنشر في المجلة تنشر بموجب المشاع الإبداعي بحقوق النشر التي تملكها مجلة الآداب والعلوم الاجتماعية. ويتم إصدار كل المقالات بموجب ترخيص الوصول المفتوح المشاع الإبداعي CC BY NC 4.0.
للاطلاع على نسخة من هذا الترخيص، يمكنكم زيارة الموقع الموالي :
http://creativecommons.org/licenses/by/4.0/
إن هذا الترخيص يسمح بإعادة استخدام المواد البحثية المفتوحة الوصول إلى الحد الأقصى. وبالتالي، فإن المعنيين بالاستفادة أحرار في نسخ ونقل وتوزيع وتكييف (إعادة خلط) المساهمات المنشورة في هذه المجلة، وهذا حتى لأغراض تجارية؛ بشرط أن يتم نسب المساهمات المستخدمة من طرفهم إلى مؤلفي هذه المساهمات، وهذا وفقًا لطريقة من الطرق المعترف بها في كتابة المراجع.

© المؤلف (المؤلفون) 2023