مجلة جامعة دمشق للعلوم الهندسية
المجلد 39- العدد 3 (2023) 177-193

Damascus university Journal for engineering sciences
Vol 39 No.3 (2023): 177-193

إيجاد أنطولوجيا عربية بالتقابل بين شبكة الكلمات العربية والموسوعة العربية

سندس محمد سعيد الحلبي[1]

أ.م.د. محمد مازن المحايري[2] **أ.د. فيصل العباس**[3]

[1] مهندسة إلكترون حاصلة على ماجستير في الحاسبات والتحكم من جامعة دمشق.

[2] أستاذ مساعد في قسم هندسة الحواسيب والأتمتة – كلية الهندسة الميكانيكية والكهربائية – جامعة دمشق.

[3] أستاذ دكتور ورئيس الجامعة الدولية الخاصة للعلوم والتكنولوجيا– رئيس شعبة الكهرباء والحاسوب في الموسوعة العربية.

الملخص

هناك اهتمام كبير في هذه السنوات في الأنطولوجيا Ontology كأداة من أجل تمثيل المعارف. يعتبر توليد الأنطولوجيا اليوم من الأبحاث الأساسية ذات أهمية في تنفيذ الكثير من التطبيقات مثل البحث الدلالي (Benassi *et al.*, 2004) والتلخيص الآلي (Carenini *et al.*, 2008) وغيرها. وقد طُوِّر العديد من المشاريع البحثية والنماذج المخبرية والأدوات من أجل هذا الغرض. إلا أن توليد أنطولوجيا عربية مازال يتطلب جهوداً بحثيةً كبيرة. ومن أكثر الأنطولوجيات العربية استخداماً شبكة الكلمات العربية Arabic Word Net (AWN) والتي وُلِدَت بترجمة شبكة الكلمات الإنكليزية إلى العربية، والمنتشرة على نحو واسع كأنطولوجيا معجمية تُستخدَم في الأبحاث الحالية بسبب إتاحتها مجاناً وتغطيتها الواسعة. ولكن على الرغم من انتشار استخدامها، إلا أنها تفتقر إلى بعض المعلومات العربية المفيدة لمعالجة اللغة. لذا طُوِّر في هذا البحث إطار لتعلّم أنطولوجيا عربية يأخذ بعين الاعتبار تحديات اللغة العربية واستخدام مصادر معرفية عربية، واستخدمت خوارزمية لتعلّم الأنطولوجيا باعتماد تقنية التقابل Mapping بين شبكة الكلمات العربية وخصائص مقالات الموسوعة العربية Arabic Encyclopedia التي تعتبر مصدر معلومات عربي غني في كافة مجالات المعرفة الإنسانية. وبناءً عليه تم إيجاد أنطولوجيا عربية جمعت بين شمولية شبكة الكلمات العربية المترجمة وبين جودة مفاهيم الموسوعة العربية ومصطلحاتها وعلاقاتها العربية. بينت النتائج أنه أُضيف عدد جيد من المفاهيم والمصطلحات العربية والعلاقات الدلالية إلى إنطولوجيا شبكة الكلمات العربية وذات دقة جيدة، وعليه يمكن اعتبار الانطولوجيا المولدة أنطولوجيا عربية أساسية Core Arabic ontology يعاد استخدامها في التطبيقات العربية الدلالية، ويمكن إستكمالها للحصول على أنطولوجيا عربية فعالة.

الكلمات المفتاحية: أنطولوجيا عربية، شبكة الكلمات العربية، تعلّم الأنطولوجيا، تقنية التقابل، الموسوعة العربية، الوب الدلالي.

تاريخ الإيداع: 2022/4/6
تاريخ القبول: 2022/10/10

حقوق النشر: جامعة دمشق –سورية، يحتفظ المؤلفون بحقوق النشر بموجب الترخيص **CC BY-NC-SA 04**

Arabic Ontology Generation By Mapping Between Arabic Wordnet With Arabic Encyclopedia

Sundos M. Said Al-Halabi[1]
Dr. Mohamad Mazen Almahayri[2] Dr. Fiasal Al-Abbas [3]

[1]An electronic engineer - Master's degree in computers and control - Damascus University.
[2]Professor at the Faculty of Computer Engineering at Damascus University
[3]Professor, and President of the International University for Science and Technology (IUST)- Head of the Electricity & Computer Department in the Arabic Encyclopedia

Received: 6/4/2022
Accepted: 10/10/2022

Copyright: Damascus University- Syria, The authors retain the copyright under a CC BY- NC-SA

ABSTRACT

There is great interest in the ontology as a tool for knowledge representation. Today, ontology learning, is considered an important research. That is used in many applications such as: semantic search (Benassi *et al.*, 2004) and automatic summarization (Carenini *et al.*, 2008) and others. There were many researches, projects, models, and tools for this purpose. However, generating an Arab ontology still requires significant research efforts. One of the most widely used Arabic ontologies is the Arabic Word Net (AWN), which was produced by translating the English Word Net into Arabic, and it is used as a lexical ontology in current researches, because of its free availability and coverage. Despite its widespread, it lacks some useful information for Arabic language processing.

Therefore, in this research, a framework for learning Arabic ontology was developed that takes into account the challenges of the Arabic language and the use of Arabic Knowledge sources, and an algorithm was adopted to learn ontology using the mapping technique between the AWN and the features of the Arabic Encyclopedia articles, which is a rich Arabic source of information in all fields of human knowledge. Accordingly, an Arabic ontology was created that combined the comprehensiveness of the AWN with the quality of the Arabic encyclopedia concepts, terminology and Arabic relations. The results showed that a good number of Arabic concepts and terms and semantic relations were added to AWN with good Precision, and therefore the generated ontology can be considered a core Arabic ontology, that can be reused in Arabic semantic applications, and completed to obtain an effective Arabic ontology.

Keywords: Arabic Ontology, Arabic Wordnet, Ontology Learning, Mapping Technology, Arabic Encyclopedia, Semantic Web.

1- المقدمة

تعتبر الأنطولوجيا أحد أهم مكونات الوب الدلالي Semantic web وحجر الأساس لعمله، وتُعرَف على أنها طريقة لتمثيل المفاهيم من حولنا عن طريق الربط بينها بعلاقات دلالية، مما يساعد على فهم أوسع للمفاهيم المختلفة. بهذه الطريقة يمكن جعل الحاسوب الآلي يصل لمستوى إدراك المعاني قريب من فهم الإنسان وإدراكه.

الأنطولوجيا هي" *توصيف صوري صريح* *formal* explicit specification للمفاهيم التشاركية shared conceptualization " (Gómez-Pèrez et al., 2004). تشير المفاهيم Conceptualization إلى النماذج المجردة لبعض الظواهر في العالم بتحديد المفاهيم المرتبطة بتلك الظاهرة. ويُقصد بكلمة الصريح Explicit هو تعريف المفاهيم المستخدمة ومحدداتها على نحوٍ صريح. وتشير كلمة صوري Formal إلى تمثيل الأنطولوجيا بطريقة صورية يمكن قراءتها ومعالجتها حاسوبياً. وترفض التشاركية Shared أن تكون الأنطولوجيا خاصة أو ضمنية ببعض الأفراد وإنما معرفة عامة وتشاركية للمجموعة.

تتألف الأنطولوجيا من جانب من: المفاهيم concepts، والعلاقات التصنيفية taxonomic relationships التي تحدد العلاقات الهرمية بين المفاهيم، والعلاقات غير التصنيفية non-taxonomic relationships بينهم، وتتألف من جانب آخر من حالات المفاهيم concept instances ومحدداتها assertions.

يمكن تمثيل الأنطولوجيا باستخدام اللغات الصورية مثل الـ XML، وRDF، وOWL.

تعلُّم الأنطولوجية Ontology learning هو جزء من استخلاص المعلومات. الهدف منه هو استخلاص المفاهيم الموائمة وعلاقاتها آلياً أو شبه آلياً من النصوص المعطاة أو أي نوع من مجموعات المعطيات لتشكيل الأنطولوجيا.

هناك تجارب ومشاريع وأبحاث كثيرة في مجال تعلّم الأنطولوجيا، ولكنها تحتاج في اللغة العربية إلى جهود وأبحاث حثيثة. لذا كان هدفنا في هذا البحث السير خطوة باتجاه توليد أنطولوجيا عربية وذلك باستخدام تقنية التقابل بين شبكة الكلمات العربية وخصائص الموسوعة العربية، للاستفادة من شمولية شبكة الكلمات العربية المترجمة وجودة الخصائص العربية في الموسوعة العربية.

وفيما يلي نعرض مقدمة نظرية عن تعلّم الأنطولوجيا وتعريف شبكة الكلمات العربية والموسوعة العربية المستخدمة في توليد الانطولوجيا، كما نلخص أهم الدراسات والأبحاث المستخدمة في توليد الأنطولوجيا العربية، ونرسم إطار لتعلم الأنطولوجيا العربية، ونسلط الضوء على تقنية التقابل في توليد الأنطولوجيا المستخدمة في بحثنا والتجارب والنتائج.

2- تعلّم الأنطولوجيا

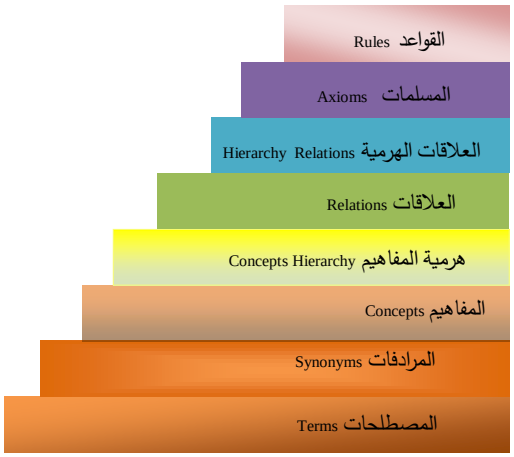
تعلّم الأنطولوجيا هو عملية التوليد الآلي أو الشبه الآلي للأنطولوجيا، متضمناً استخراج المصطلحات المرتبطة بالمجال والعلاقات بين المفاهيم التي تمثلها هذه المصطلحات من مجموعة نصية مكتوبة بلغة طبيعية، وترميزها بلغة الأنطولوجيا لسهولة استرجاعهم. وقد جاءت ضرورة تعلّم الأنطولوجيا من عدة عوامل: منها كلفة بناء الأنطولوجيا المرتفعة، والتغير المستمر في العلوم والمعارف.

صنف الباحثون في (Drumond *et al.*, 2008) تعلّم الأنطولوجيا حسب نوع معطيات الدخل إلى:

- تعلّم الأنطولوجيا من المعطيات المهيكلة structured data: تولد الإجراءات في هذه الطريقة أجزاء الأنطولوجيا المستخلصة باستخدام المعطيات المهيكلة مثل قواعد المعطيات وقواعد المعرفة وأنطولوجيات موجودة مسبقاً.

- تعلّم الأنطولوجيا من المعطيات شبه المهيكلة semi-structured data: وكمثال على هذه المعطيات المعاجم، والـ Word Net، ووثائق XML و HTML، وغيرها.
- تعلّم الأنطولوجيا من المعطيات غير المهيكلة unstructured Data: تتكون الوثائق غير المهيكلة من نصوص بلغات طبيعية مثل وثائق الـ Word أو الـ PDF أو صفحات الوب. وتستخدم عادةً طرق لغوية أو إحصائية لاستخراج المعرفة من هذه المصادر.

رتبت المهام الأساسية المتعددة والمتعلقة بتعلّم الأنطولوجية في رسم بياني على شكل طبقات كعكة في (Cimiano *et al.*, 2006) وهي موضحة في الشكل (1). تُرتب الطبقات فيها فوق بعضها باعتبار أن النتائج في الطبقات الأدنى تطبق كدخل للطبقات الأعلى منها. الطبقتان السفليتان في الكعكة تخص المستوى المفرداتي lexical level في تعلّم الأنطولوجية. مهمة هذا الجزء تحديد المصطلحات وتجميعها إلى مصطلحات مترادفة يمكن استخدامها كقاعدة لتشكيل المفاهيم. وفي المستويات الأعلى من طبقات الكعكة نجد الطبقات التي تخص مهام تعلّم هرمية المفاهيم، والعلاقات، وهرمية العلاقات بالإضافة إلى القواعد rules والمسلمات axioms.



الشكل (1) طبقات كعكة تعلّم الأنطولوجية

استخدمت في العقود الماضية عدة تقنيات من أجل توليد الانطولوجيا (Asim *et al.*, 2018) منها: معالجة اللغات الطبيعية والتعلم الآلي واسترجاع المعلومات والتنقيب عن المعلومات. وفرت تقنيات التنقيب عن المعلومات واسترجاع المعلومات والتعلم الآلي آلية إحصائية لاستخلاص المصطلحات والمفاهيم وارتباطاتهم. ومن جانب آخر لعبت تقنيات معالجة اللغات الطبيعية دوراً في معظم مراحل تعلم الأنطولوجيا.

كما استخدمت تقنيات التقابل Mapping أو الدمج merging في العديد من التطبيقات مثل: ((Szymanski *et al.*, 2012) (Fernando *et al.*, 2009)، (Soor, 2015)، (John, 2017)، (Bao *et al.*, 2004)، (Ehrig *et al.*, 2005)، (Juzaddin *et al.*, 2015)، (Ruiz-Casado *et al.*, 2005)، (Suchanek *et al.*, 2008)، (Ponzetto *et al.*, 2010)) من أجل: دمج عدة أنطولوجيات، أو توسيع الأنطولوجيا بتقابل مفاهيمها مع مصادر معرفية شبه مهيكلة مثل الموسوعات وذلك إما بربط عناوين أو اختصاصات أو خصائص مقالات الموسوعات بالمفاهيم في شبكة الكلمات.

لا يوجد قياسات كثيرة ضابطة لتقييم تعلّم الأنطولوجيا كون علم تعلّم الأنطولوجيا حديث نوعاً ما. كما أن نتائج هذا النوع

من التعلّم ليست قائمة بسيطة بتصنيفات ثنائية وإنما لها بنية معقدة . فكما صرح الباحث في (Smith, 2004) بأن هناك احتمالات متعددة للتجريد في مجال واحد تختلف باختلاف استخدامها واختلاف مجموعات الأشخاص. فإذا لم تقارن نتائج الخوارزمية المطبقة في تعلّم الأنطولوجيا مع أنطولوجيا جيدة مبنية يدوياً فلا يمكن تقييم جودتها. ولأن الأنطولوجيا ليست بحد ذاتها هي موضع الاهتمام وإنما التطبيقات التي تستخدم فيها كمصادر لتحسين الأداء، فهناك عدة طرق يمكن تطبيقها للتقييم الآلي للأنطولوجيات (Biemann *et al.*, 2005) تتضمن: التقييم باستخدام معيار الذهب a Gold Standard وذلك بمقارنة الأنطولوجيا المتولدة مع أنطولوجيا المرجعية، التقييم باعتماد التطبيق Application-based وذلك بقياس مدى مساعدة الأنطولوجيا المتولدة في تحسين نتائج التطبيق، والتقييم باعتماد البيانات Data driven .

3- البحوث والدراسات السابقة لتوليد انطولوجيا عربية

يمكن تلخيص أهم المشاريع والدراسات السابقة لتوليد انطولوجيا عربية بما يلي:

اعتمد في توليد أنطولوجيا عربية إما على ترجمة أنطولوجيات أجنبية، كما في مشروع (AlBarhamtoshy *et al.*, 2009) حيث قام الباحثون فيه بتصميم وتنفيذ شبكة كلمات عربية Arabic Word Net عن طريق ترجمة شبكة الكلمات الإنكليزية Word Net إلى العربية باستخدام قاموس متعدد اللغات. كما تم أيضاً توليد شبكة الكلمات العربية في مشروع(Rodríguez H. 2008) بتوسيع شبكة الكلمات العربية بشكل شبه آلي باستخدام قواعد صرفية وقواعدية وتطبيق استدلال بايزن على شبكة الكلمات الانكليزية.

أو بتوسيع أنطولوجيات عربية مترجمة اعتماداً على مصادر معرفية مترجمة، كما في مشروع (Alkhalifa *et al.*, 2008) حيث تم توسيع شبكة الكلمات العربية آلياً بإضافة أسماء كيانات عربية من عناوين مقالات موسوعة الويكيبديا، وفي مشروع (Juzaddin *et al.*, 2015) حيث تم توسيع شبكة الكلمات العربية بالتقابل بين مرادفات شبكة الكلمات العربية وخصائص مقالات موسوعة الويكيبيديا المترجمة إلى العربية. واستخدمت الأنطولوجيا اللغوية Gold كنواة في مشروع أنطولوجيا الخليل (Aliane *et al.*, 2009) لبناء البنية التحتية لأنطولوجيا عربية مركزية. كما قام الباحثون في مشروع (Mazari *et al.*, 2012) ببناء أنطولوجيا في مجالات لغوية تخص اللغة العربية معتمدين المفاهيم العامة في أنطولوجيا Gold ومجموعة خصائص نصوص عربية واستخدام تقنيات إحصائية تعتمد على توارد العبارات في تلك النصوص، كما اعتمد الترافق لربط المفاهيم المستخلصة بالأنطولوجيا بعلاقات هرمية وغير هرمية.

كما طور الباحثون مشروع الأنطولوجيا العربية بجامعة بيرزيت (Jarrar, 2011) (Jarrar M., 2021) أنطولوجيا عربية بتوسيع شبكة الكلمات العربية اعتماداً على المعاجم باستخراج المفاهيم العربية ومقابلتها مع شبكة الكلمات العربية.

كما استخدمت طرق إحصائية لإنشاء أنطولوجيا عربية متخصصة بمجال معين، كما في مشروع (Alsudias L. *et al.*, 2020) حيث وُلدت أنطولوجيا عربية طبية متخصصة بالأمراض المعدية حيث استخرجت المصطلحات من مستندات طبية في هذا المجال آلياً وجرى مكاملتها مع معاجم علمية. وفي مشروع (Mohammed *et al.*, 2014) حيث تم توليد علاقات دلالية آلياً من نصوص عربية باعتماد الأنماط القواعدية. كما وُلِدَت في مشروع (Belkredim *et al.*, 2008) أنطولوجيا الديار DEAR-ONTO وهي أنطولوجيا عربية اشتقاقية بينية، استُخدِمت فيها الأفعال السماعية كجذور وحددت الأصناف المختلفة المشكلة للكلمات المشتقة منها.

كما كانت هناك عدة بحوث ومشاريع لتوليد أنطولوجيات للقرآن الكريم مثل شبكة الكلمات العربية القرآنية Quranic Arabic WordNet (QAWN) في (AlMaayah *et al.*, 2016) وذلك بتوليد المرادفات آلياً مستندين على قواميس عربية. وأنطولوجيا القصص القرآنية باستخدام تقنية التقابل في مشروع (Al-Salhi R.Y. 2022). وفي مشروع أنطولوجيا الأزهري Azhary (Ishkewy *et al.*, 2014) تم إيجاد انطولوجيا عربية معجمية وذلك باستخراج الكلمات والمفاهيم العربية آلياً من نصوص القرآن الكريم وربطها يدوياً بعلاقات استناداً إلى معاجم عربية عربية.

لذا نجد بأن هذه المشاريع والبحوث التي اقتصرت على توليد الأنطولوجيا العربية بالترجمة أو بتوسيع أنطولوجيا مترجمة من مصادر عربية مترجمة مثل موسوعة الويكيبيديا. كما استخدمت طرقاً احصائية لتوليد بعض الأنطولوجيات العربية المتخصصة يدوياً في مجال معين اعتماداً على معاجم وقواميس ومصادر معلومات نصية عربية في ذلك المجال.

4- شبكة الكلمات العربية:

شبكة الكلمات هي مستودع معرفي هرمي البنية غني بالعلاقات المعجمية الصريحة أُنشئ من قبل خبراء لغويين. وهو إنتاج مشروع بحثي تم تطويره في جامعة Princeton للحصول على معارف معجمية للغة الأم وحفظه في قاموس يمكن قراءته من قبل الآلة (Fellbaum ,1998). تحتوي كلمات مصنَّفة إلى أسماء وأفعال وصفات وظروف، وكلمات أخرى من نوع ضمائر وحروف جر وعطف. تُنظَّم الكلمات ضمن شبكة الكلمات في مجموعات ترادف (المفاهيم) تسمى Synsets وهي تشكل البنية الأساسية للشبكة. يمكن لكلمة واحدة أن تحمل أكثر من معنى، يعطى لكل معنى فيها عنصر يسمى معنى الكلمة word sense. تحتوي مجموعة المرادفات على معاني للمساعدة في وصفها. لكل مجموعة مرادفات شرح gloss يوصف معنى الكلمة. كما ترتبط معظم المجموعات بأمثلة استخدام.

ترتبط المعاني والمفاهيم بعلاقات متنوعة. تسمى العلاقات التي تربط بين المعاني العلاقات الدلالية Semantic Relations، بينما تسمى تلك التي تربط المفاهيم علاقات معجمية Lexical Relations. ترتبط الأسماء بعلاقات دلالية مثل: الجزء/التضمين Hyponym/Hypernym، والانتماء/الاشتمال Meronym/Holonym، وجزء من/لها جزء Meronym/Holonym، ومستوى/له مستوى Meronym/Holonym. تشكل هذه العلاقات تصنيفات ذات بنية هرمية. تنتمي بعض المجموعات إلى تصنيف واحد بينما ينتمي البعض الآخر إلى أكثر من تصنيف. كما يوجد علاقات معجمية أخرى مثل المضاد antonyms، والمقاد derived-from. تكون العلاقات المعجمية بين معاني الكلمات وليس بالضرورة بين المفاهيم.

تُمثَّل كلمات شبكة الكلمات بتنسيق خاص ولكل كلمة نوع الكلام Part-of-Speech (POS) ورقم المعنى. ويوجد أربعة أنواع للكلام في الشبكة هي: n للأسماء، وv للأفعال، وa للصفات، وr للضمائر.

وقد تُرجمت شبكة الكلمات إلى عدة لغات من ضمنها العربية فشكلت شبكة الكلمات العربية Arabic Word Net AWN وهي أنطولوجيا عربية معجمية استخدمت على نحو واسع، إلا أنها تفتقر إلى بعض المفاهيم والعلاقات العربية ولها عدة تحديات (Black *et al.*, 2006) كونها تشكلت عن طريق الترجمة.

5- تعريف الموسوعة العربية:

تعتبر الموسوعة العربية أكبر الموسوعات العربية المتوفرة ورقياً وإلكترونياً، مكتوبة بأقلام عربية بمصطلحات وتعابير عربية وهي ليست ترجمة لثقافة ومصطلحات غريبة عن اللغة العربية. وهي من تأليف باحثين عرب تم تأليفها في هيئة

الموسوعة العربية في الجمهورية العربية السورية، صدرت نسختها الإلكترونية على موقع هيئة الموسوعة العربية منذ عام 2010 ومتاحة مجاناً للعموم. تتألف من حوالي 15000 بحثاً في مجالات المعرفة الإنسانية جميعها التي تتضمن: الحضارة العربية، والعلوم الإنسانية، والآداب واللغات، والعلوم الطبية، والعلوم البحتة، والعلوم التطبيقية، والعلوم القانونية، والتربية والفنون.

تتألف كل مقالة في الموسوعة من نص يصف موضوع المقالة مع روابط للمواضيع ذات الصلة. الهدف من الروابط هو تزويد القرّاء بمعلومات إضافية مرتبطة بالموضوع. إضافةً للنص الذي يصف المقالة فهناك مجموعة مصطلحات تُشكّل عنوان فريد للمقالة. وعندما تناقش مقالتان أو أكثر موضوعاً بكلمة تحمل أكثر من معنى يتم إرفاق العنوان عادة بمصطلح وصفي للتمييز بين المادتين.

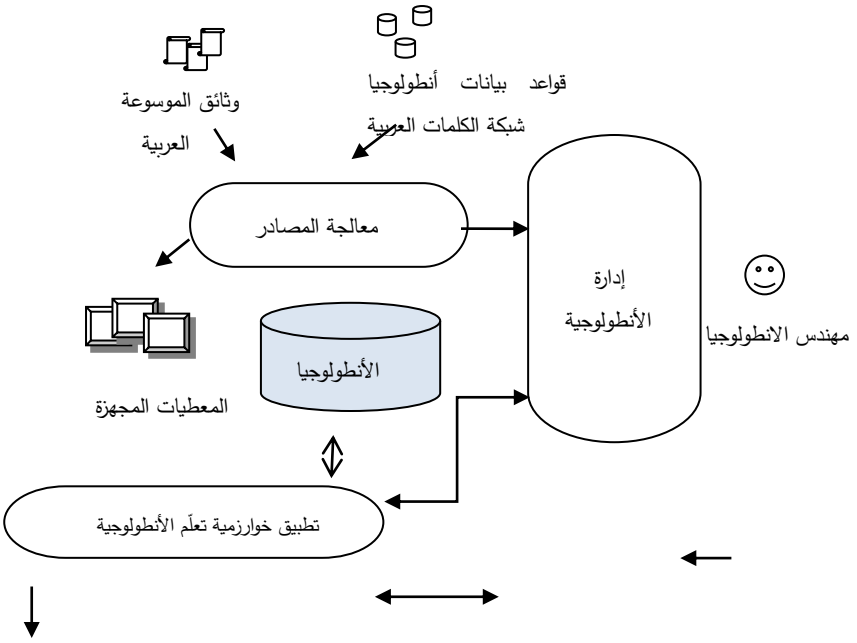
توصف المقالات الموجودة في الموسوعة العربية، كما في الموسوعات الأخرى عدة مواضيع. يمكن اعتبار كل مقالة مفهوم يرتبط بمحتوى نص المقالة الذي يصف موضوعها الأساسي.

6- توليد الأنطولوجيا العربية بالتقابل بين شبكة الكلمات العربية والموسوعة العربية

نتيجة دراسة المشاريع السابقة في مجال توليد أنطولوجيا عربية تبين لنا أنه لا يوجد أنطولوجيا عربية تضم المفاهيم العربية ومجموعة ترادفات المصطلحات العربية وعلاقاتها الدلالية إلا في بعض المشاريع البحثية غير المكتملة والتي تعتمد في توليدها على الترجمة سواء من مصادر معرفية مثل موسوعة الويكيبيديا أو من أنطولوجيا مثل شبكة الكلمات الإنكليزية أو يتوسيع الانطولوجيا المترجمة اعتماداً على مصادر معرفية مترجمة إلى العربية. لذا حاولنا في هذا البحث إيجاد أنطولوجيا عربية اعتماداً على توسيع انطولوجيا الكلمات العربية وإغنائها بالاستعانة بخصائص الموسوعة العربية التي تعتبر أفضل مصدر معرفي عربي غني بالمعلومات والمفاهيم العربية في كافة المجالات وذلك باستخدام تقنية التقابل.

إن عملية تعلّم الأنطولوجيا لا تختلف بشكل جوهري عن عملية التنقيب عن المعطيات وذلك بمراحل فهم المعطيات، وتجهيزها، ونمذجتها، وتقييمها وتوظيفها.

لذا فإن المكونات الأساسية لإطار تعلم الإنطولوجيا المقترح كما هو مبين في الشكل (2) هي: مكون إدارة الأنطولوجيا، ومكون معالجة المصادر، ومكون خوارزمية تعلّم الأنطولوجيا.



الشكل (2) إطار تعلم الأنطولوجيا العربية بالتقابل بين شبكة الكلمات العربية والموسوعة العربية

وسوف نصف هذه المكونات فيما يلي:

6-1 مكون إدارة الأنطولوجيا:

يستخدم هذا المكون من قبل مهندس الأنطولوجيا لمعالجة الأنطولوجيا. تقدم أدوات إدارة الأنطولوجيا الخدمات التالية: واجهة تخاطبية لتصفح الأنطولوجيا وتعديلها وتقييمها.

6-2 مكون معالجة المصادر المعرفية:

يحتوي هذا المكون تقنيات من أجل استكشاف معطيات الدخل واستيرادها وتحليلها وتحويلها. الهدف منه هو توليد مجموعة معطيات معالجة ومجهَّزة كدخل لمكون خوارزمية تعلّم الأنطولوجيا. يتضمن هذا المكون أداتين، أداة لتوليد خصائص الموسوعة العربية وأداة لتجهيز بيانات شبكة الكلمات العربية.

6-2-1 أداة توليد خصائص الموسوعة العربية:

الهدف من توليد خصائص الموسوعة (Fawcett *et al.*, 1993) هو إثراء تمثيل المستند النصي من خلال تعزيزه بخصائص مُستخرَجة منها. تتضمن هذه الخصائص المفاهيم المُكتشَفة وعلاقاتها ضمن المستند. يُمثَّل كل مفهوم مُكتشَف بشعاع عناصره المفاهيم كلها مع درجة ترابط بينها وبين المفهوم الرئيسي. وتُلغَى المفاهيم التي درجة ارتباطها ضعيفة من قائمة مفاهيم الشعاع.

تتضمن هذه الأداة قسمين، قسم لاستخراج الخصائص وتثقيلها، وقسم لحساب العلاقات بين المصطلحات.

أولاً– استخراج الخصاص وتثقيلها:

تتضمن هذه المرحلة المعالجة الأولية لنصوص مقالات الموسوعة وفهرستها فهرسة آلية لاستخراج المصطلحات وتثقيلها اعتماداً على تواردها في نصوص الموسوعة وفقاً للخطوات المبينة في الشكل (3) التقطيع Tokenization، والتقييس Normalization، وإزالة كلمات التوقف Removing Stopword، والتجريد Stemming، وتثقيل الكلمات المفتاحية Weighting of keywords.

الشكل (3) خطوات استخراج خصائص الموسوعة وتثقيلها

التقطيع: يُحوَّل الشريط المحرفي (المستند) إلى مجموعة مصطلحات اعتماداً على محددات مثل المسافة البيضاء والفاصلة المنقوطة والاقتباسات، كما تولد عبارات بتقسيم الشريط المحرفي إلى عبارات بكلمتين. من خلال إجرائية مستندة إلى خوارزمية الأداة AraNLP مع بعض التعديلات (Althobaiti *et al.*, 2014)

التقييس: في اللغة العربية، قد تظهر بعض الحروف العربية بأشكال مختلفة، وتستخدم حروف بدلاً من حروف لأن أشكالها متشابهة. كما يستخدم الكتّاب علامات التشكيل في نصوصهم، التي يمكن أن تؤدي إلى اختلافات للمصطلح نفسه، وبالتالي تؤثر على حساب بعض الخصائص مثل التردد (TF). لذلك فإن عملية التقييس مطلوبة لتوحيد الأشكال المختلفة للحرف نفسه لتجنب مثل هذه الالتباسات. استُخدمَت إجرائية التقييس المقترحة الخوارزمية المطبقة في أداة AraNLP للقيام بالمهام التالية (Althobaiti *et al.*, 2014): إزالة الأحرف غير العربية مثل الرموز وعلامات الترقيم الخاصة، وإزالة علامات التشكيل، واستبدال "ا" بـ "أ إ آ"، و"ي" بـ "ى"، و"ه" بـ "ة"، و"و" بـ "ؤ"، و"ي" بـ "ئ"، وإزالة التطويل.

إزالة كلمات التوقف: كلمات التوقف (مثل الضمائر وحروف الجر وحروف العطف وما إلى ذلك) هي كلمات غير مهمة تظهر كثيراً في المستندات لتكوين الجمل (Kanan *et al.*, 2004). نظراً لأن هذه الكلمات ليست مفيدة (لا تضيف معلومات)، فيمكن حذفها من الجمل دون التأثير على المحتوى الأساسي للجملة. في الواقع هناك العديد من قوائم التوقف المعتمدة في تطبيقات التي تحتاج إزالتها، اعتُمِد في هذا البحث على قائمة التوقف العامة المستخدمة في أداة AraNLP، وهي القائمة المعتمدة في معظم التطبيقات مع بعض الإضافات عليها.

التجريد: التجريد بالعربية عملية شاقة ودقيقة، لأن اللغة العربية هي لغة صرفية واشتقاقية بحتة، مما يعني أن الكلمات العربية يمكن أن يكون لها العديد من الأشكال المختلفة ولكنها تشترك في المعنى المجرد نفسه. إن عملية التجريد هي عملية إزالة بعض أو كل اللواحق (مثل البادئات، واللواحق، واللواحق اللاحقة) من الكلمة وذلك بتحويل الأشكال والمشتقات المختلفة للكلمة إلى شكل موحد مثل الجذر أو الجذع. اعتمد في هذا البحث طريقة التجريد بحذف السوابق واللواحق باستخدام إجرائية تعتمد الخوارزمية والبيانات المطبقة في أداة خوجة.

تثقيل الكلمات المفتاحية: تثقل المصطلحات (الكلمات المفتاحية) الواردة في مستند معين لمعرفة أهميتها فيه، استخدمت المعادلة رقم (1) وهي علاقة التثقيل المشهورة من أجل حساب وزن المصطلحات والتي تم تقييسها بضرب وجمع العدد '0.1' الذي تم اختياره تجريبياً.

$$w = 0.1 + 0.1\left(\frac{t_{fi}}{max_{t_{fi}}}\right)\log(\frac{N}{n_i}) \qquad (1)$$

حيث: t_{fi} تردد المصطلح في الجملة، و N عدد الجمل في المستند، و n_i عدد الجمل التي تحتوي المصطلح i.

ثانياً– أداة حساب العلاقات بين الخصائص:

استخدمت هذه الأداة تقنيات إحصائية لاكتشاف وحساب العلاقات بين المصطلحات والمفاهيم المستخرجة من الموسوعة العربية، والعلاقات التي تولدت منها هي علاقة الترافق Co-Occurance Relations، وعلاقة التشابه Similarity Relations.

علاقة الترافق: حسبت هذه العلاقة اعتماداً على ترافق مصطلحين معاً وذلك باستخدام المعادلة رقم

(2) (Van Rijsbergen 1976) والتي تربط بين مصطلحين اعتماداً على تواردهما المشترك.

$$CoOccurance(t_u, t_v) = \frac{CoOccurance\ Frequency}{\sqrt{Frequency(t_u) * Frequency(t_v)}} \quad (2)$$

حيث t_u وt_v المصطلحين المرتبطين بعلاقة ترافق، و CoOccuranceFrequency هو تردد حدوث المصطلحين t_u وt_v معاً في مجموعة مستندات الموسوعة. و$frequency(t_u)$ هو تردد حدوث المصطلح t_u في المجموعة، و$frequency(t_v)$ هو تردد حدوث المصطلح t_v في المجموعة.

علاقة التشابه: ترتبط المصطلحات مع بعضها باعتماد تواردها في الموسوعة، يمكن حساب هذا النوع من العلاقات باستخدام علاقة التجيب بين أشعة المصطلحات، حيث يُمثَّل كل مصطلح بشعاع مركباته أوزان هذا المصطلح في مستندات الموسوعة، وذلك وفقاً للمعادلة رقم (3).

$$Sim_{t_u,t_v} = \frac{\sum_{i=1}^{p} w_{ui} w_{vi}}{\sqrt{\sum_{i=1}^{p} w_{ui}^2} \sqrt{\sum_{i=1}^{p} w_{vi}^2}} \quad (3)$$

حيث: D مجموعة المستندات في الموسوعة العربية، و P عدد المستندات في D، وw_{ui} وزن المصطلح u في المستند i، وW_{vi} وزن المصطلح v في المستند i.

6-2-2 أداة تجهيز بيانات شبكة الكلمات العربية:

استخدمت أداة تجهيز بيانات شبكة الكلمات العربية لفهرسة مفاهيم ومصطلحات الشبكة التي تتكون من 11270 مفهوماً و23496 مصطلحاً. وذلك بتطبيق طريقة فهرسة مصطلحات الموسوعة العربية نفسها، من أجل إتاحة إمكانية مقارنتها مع مفاهيم ومصطلحات الموسوعة العربية في عملية التقابل المقترحة، وفق المراحل التالية: التقييس، وإزالة كلمات التوقف، والتجريد.

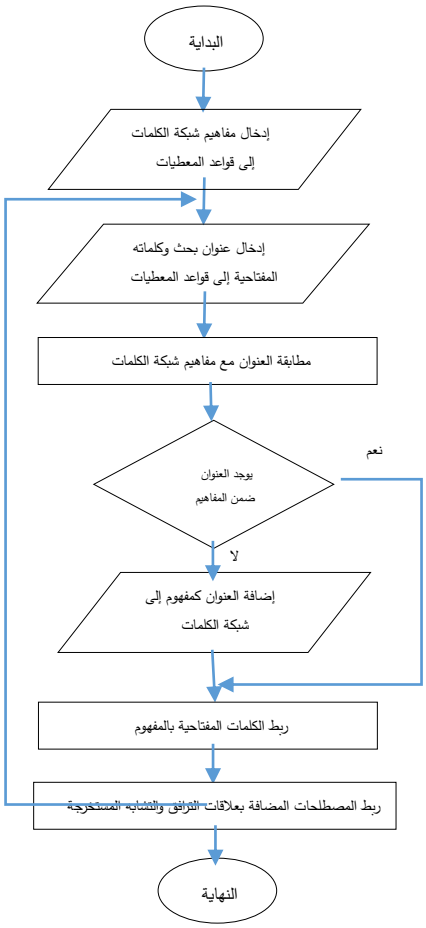
6-3 مكون تطبيق خوارزمية تعلّم الأنطولوجيا:

يهدف هذا المكون إلى توليد الأنطولوجيا العربية اعتماداً على خوارزمية تعلم الأنطولوجيا المقترحة والتي تهدف إلى إيجاد أنطولوجيا عربية باستخدام تقنية التقابل لتوسيع أنطولوجيا شبكة الكلمات العربية، وذلك بإغناء الشبكة بإضافة مفاهيم ومصطلحات الموسوعة العربية إليها وربطها بالعلاقات المستخرجة وذلك اعتماداً على البيانات الناتجة عن مكون معالجة المصادر. وطبقت خوارزمية التعلم الموضحة في المخطط في الشكل رقم (4) وفقاً لما يلي:

1- مطابقة مفاهيم شبكة الكلمات مع عناوين مقالات الموسوعة، وإضافة عناوين المقالات غير الموجودة في الشبكة كمفاهيم جديدة باستثناء أسماء الأعلام.

2- إضافة المصطلحات الهامة في مستندات الموسوعة (التي لها ثقل يتجاوز 0.69 في مستندات الموسوعة) إلى الشبكة بضمها إلى المفهوم المناسب في الشبكة.

3- ربط المصطلحات والمفاهيم المضافة إلى الشبكة بالعلاقات المستخرجة وهي علاقات الترافق والتشابة (التي لها ثقل يتجاوز 0.5).



الشكل (4) خوارزمية تعلم الأنطولوجيا

7- التجارب والنتائج:

تم تطبيق إطار تعلم الأنطولوجيا بمكوناته المقترحة وذلك ببناء برامج لكل مكون كما استخدمت قواعد البيانات أوراكل وأدوات تطويرها من أجل معالجة البيانات وإدارتها.

أولاً- استخدم البرنامج المرتبط **بمكون معالجة المصادر المعرفية** من أجل توليد خصائص الموسوعة العربية وتجهيز بيانات شبكة الكلمات العربية وفقاً لما يلي:

- استخدمت الموسوعة العربية كمصدر معرفي غني يضم 15000 مقالة مكتوبة بلغة عربية صحيحة في مجالات المعرفة الإنسانية كافة. فُهرِست هذه النصوص فهرسة آلية بتطبيق أداة توليد خصائص الموسوعة العربية التي تتضمن مجموعة من العمليات وهي: التقطيع، والتقييس، وإزالة كلمات التوقف، والتجريد، وتثقيل الكلمات المفتاحية. كما ولدت علاقات التشابه والترافق بين المصطلحات المولدة باستخدام أداة حساب العلاقات بين الخصائص.
- جهزت بيانات شبكة الكلمات العربية AWN باستخدام أداة تجهيز بيانات شبكة الكلمات العربية والتي تتألف من 11270 مفهوماً و23496 مصطلحاً وذلك بتطبيق مراحل التجريد والتقييس وإلغاء كلمات التوقف.

ثانياً- استخدم البرنامج المرتبط **بمكون تطبيق خوارزمية التعلم** لتوسيع شبكة الكلمات العربية بإغنائها بمفاهيم ومصطلحات وعلاقات مستخرجة من الموسوعة العربية باستخدام خرج مكون معالجة المصادر المعرفية كدخل له وتم التوصل إلى:

- عدد المفاهيم المضافة إلى شبكة الكلمات 5778 مفهوماً.
- عدد المصطلحات المضافة 10212 مصطلحاً.
- عدد العلاقات المضافة بين المصطلحات من نوع ترافق: 3500 علاقة.
- عدد العلاقات المضافة بين المصطلحات من نوع تشابه: 2400 علاقة.

ثالثاً- استخدم البرنامج المرتبط بمكون **إدارة الأنطولوجيا** من أجل تحرير وتنقيح الأنطولوجيا العربية المتولدة وتقييمها.

يبين الجدول رقم (1) عينات من مجموعات المفاهيم المجردة والمصطلحات المجردة المرتبطة في الأنطولوجيا العربية. كما يبين الجدول رقم (2) عينات من علاقات التشابه والترافق بين المصطلحات المجردة في الأنطولوجيا العربية.

يبين الجدول رقم (3) احصائيات عن المفاهيم والمصطلحات والعلاقات المضافة إلى شبكة الكلمات العربية.

الجدول رقم (1) عينات من مجموعات المفاهيم والمصطلحات المرتبطة في الأنطولوجيا العربية

المصطلح المجرد المرتبط بالمفهوم	المفهوم المجرد
ابعاد سيار	ابعاد مركب
حفار هدروليك	
سع قاطر	
سيار شاحن	
سيار قاطر	
عرب متحرك	
مقطور	
نصف مقطور	
عناصر لانتانيديه	اترب نادر
غادولينيوم	
لانتانيديه	
لوتسيوم	
اتربى	اتريوم
اكسيد اتريوم	
املاح اتريوم	
خواص اتريوم	
زرنيخ اتريوم	
فسفور اتريوم	
فلور اتريوم	
كبريت اتريوم	
كربون اتريوم	
كلوريد اتريوم	
نترات اتريوم	
هدريد اتريوم	
يوديد اتريوم	
تابع تحويل	اتماتي
تحكم رقم	
تحويل منظوم	
تصميم منظوم	
محاك منظوم	
منظوم متقطع	
منظوم مستمر	
انتاج موتمت	اتمت
تطبيق اتمت	
منظوم موتمت	
اثر بلتيه	اثر بلتيه اثر طمسون
اثر طومسون	
اشعاع تشيرنكوف	اثر تشيرنكوف
تشيرنكوف	
عداد تشيرنكوف	
اجاز صح	اجاز عمل
اجر كامل	
عامل اجاز	
مجال اجتهاد	اجتهاد
مصادر اجتهاد	
اجهاض جريم	اجهاض
اجهاض دوا	
اجهاض عفو	
اجهاض متكرر	
اجهاض محرض	
تجريف	
جريم اجهاض	
حامل اجهاض	
عقوب اجهاض	
اجهز تعليق	اجهز تعليق مركب
تعليق مركب	
تعليق مرن	
جهاز تعليق	
عارض استناد	
كراس مضاجع	
مجال التو	
مجموع تعليق	
مركب اجهز	
مركب ثبات	
مركب مزنجر	
مضاجع عجل	
نظام تعليق	
هدروهواى	
هيكل مركب	
راصد	اجهز رصد فلك
كاسغران	
احباط ضعيف	احباط
سلوك عدوان	
احتشاا شامل	احتشـا قلـب امـراض شرايين اكليل
اصاب احتشاا	
اكليل سوى	
الم خناق	
الم صدر	
تصوير شرايين	
تضيق عصيد	
حال احتشاا	
خناق صدر	
خناق عاد	
دوا خناق	
دينميه دم	
رجف بطين	
صفيح عصيد	
عصيد شريان	
اجر محاكم	احداث جانح
تدبير اصلاح	
جانح سنه	
حدث جانح	
محاكم احداث	
محكم احداث	
كتاب خواتيم	اختام
نقش اختام	
احداث اختلاج	اختلاج
اعتلال دماغ	
امراض دماغ	
نقص كلسميه	
يوريميه	
اختناق جنين	اختناق
اعط اكسجين	
اكسجين نسب	
تنبيب رغامى	
حال اختناق	
اخراج مجل	اخراج صحف
اسلوب اخراج	
مخرج صحف	
كفاى عمل	ادار انتاج
مدير انتاج	
مصنع تجهيز	
احتياط اهتلاك	ادار مال
تحليل مال	
تمويل قصير	
خط مال	
عام تمويل	
عمل جارى	
ادرينالين نورادرينالين	ادرينالين
تنبيه اعصاب	
لب كظر	

نورادرينالين	
حوروس	ادفو
معبد ادفو	
ادمان	ادمان سموم
ادمان اعط	
اعتياد عضو	
امفيتامين	
ايتلاميد	
ايقاف تعاط	
تدخين	
تعاط	
تعاط حشيش	
حدوث ادمان	
حمض ليزرجيك	
سموم ادمان	
ادنبره حديث	ادنبره
ادنبره حديق	
ادنبره عصور	
ادنبره قديم	
ادنبره مدين	
تيار وشيع	ادو قياس كهرب
دار قياس	
قياس تيار	
كهرديناميك	
مقياس غلفان	
وشيع متحرك	
اذن ظاهر	اذن انسان
اذن وسط	
اذن وظيف	
عصب سمع	
عظم ركاب	
غشاا طبل	
نفير اوستاش	
نقص سمع	
وسط اذن	
وظيف اذن	

الجدول (2) علاقات التشابه والترافق من الأنطولوجيا العربية

نوع العلاقة	المصطلح الثاني	المصطلح الأول
تشابه	اببا	ابابا
تشابه	طبول	ابواق
تشابه	نقوى	ابيضاض
تشابه	مصف	اتعاب
تشابه	انتربولوجي	اتنولوجي
تشابه	مزنق	احتجاز
تشابه	فلول	احتم
تشابه	عاتي	اخمج
تشابه	نورادرينالين	ادرينالين
تشابه	بطين	اذين
تشابه	سابور	اذين
تشابه	ارثوس	ارادوس
تشابه	كاريب	اراوك
تشابه	رجاز	ارجز
تشابه	راحم	ارحم
تشابه	شجاج	ارش
تشابه	اساربع	ازعج
تشابه	اخضور	ازهوري
تشابه	اشاح	اساده
تشابه	متعالج	استبصار
تشابه	جحد	استحل
تشابه	بلمر	استر
تشابه	استركان	استراخان
تشابه	استروجين	استراديول
تشابه	استربول	استراديول
تشابه	محفوز	استشراق
تشابه	استدان	استقراض
تشابه	استنسال	استنساخ
تشابه	مرقط	اسقمري
تشابه	ارفق	اسمح
تشابه	منتفخ	اشرع
تشابه	اخرف	اصاف
تشابه	اقتل	اضرب
تشابه	معار	اعار
تشابه	اثوب	اقت
تشابه	افحاش	اقذاع
تشابه	اماجد	اكارم
تشابه	مكدس	اكسال
تشابه	حاجر	اكناف
تشابه	استجار	الطاف
تشابه	البرولين	الغليسين
تشابه	البروتيوكليكان	الغليكوبروتين
تشابه	المالتوز	الفركتوز
تشابه	الهو	اللور
تشابه	عيال	انفع
تشابه	فربح	انيس
تشابه	المسترانول	انيسترول
تشابه	شجن	انين
تشابه	حسيره	اهات
تشابه	حصاير	اوتاد
تشابه	عتبي	بستي
تشابه	بازلاا	بسله
تشابه	سبحل	بسمل
تشابه	نفتر	بهتان
تشابه	بيزانط	بيزانتيون
تشابه	ثاليوم	بيزموت
تشابه	توان	تابق
تشابه	تخشب	تابل
تشابه	تاريف	تاربث
تشابه	غراثيان	تاسار
تشابه	كورار	تاقي
تشابه	تشاكل	تباكل
تشابه	تبذير	تبذر
تشابه	استحجر	تخشب
تشابه	تسبيع	تخميس
تشابه	تيرك	ترتش
تشابه	احتيج	ترخيم
تشابه	تحمص	ترذ
تشابه	بسله	ترمس
تشابه	ديثرامب	ترنيم
تشابه	ستراتوسفير	تروبوسفير
تشابه	تعشير	تسبيع
تشابه	مماطل	تسويف
تشابه	شايب	تشويه
تشابه	انغمار	تشوك
تشابه	عصيد	تصلب
تشابه	تحجيب	تطعيم
تشابه	بلاعم	تعاضد

تشابه	اكسين	تعاير
تشابه	تشفير	تعمي
تشابه	جدانوف	تغانروغ
تشابه	بلاصر	تغلات
تشابه	مبتسر	تفحم
تشابه	ومضاني	تفرس
تشابه	كسل	تقبح
تشابه	تعتيق	تقس
تشابه	تطريق	تكتيل
تشابه	منقل	تكسر
تشابه	اذره	تلبيد
تشابه	ثرثر	تلغو
تشابه	كادميوم	تلوريوم
تشابه	غسول	تليلات
تشابه	ايحايي	تنويم
تشابه	كلكل	تنيخ
تشابه	طهرق	تهارك
تشابه	سنكيل	توان
تشابه	مسود	تودع
ترافق	اعشاب	اباد
ترافق	جماع	اباد
ترافق	مركب	ابعاد
ترافق	دم	ابيضاض
ترافق	اوكايالي	ابيوريماك
ترافق	هفايستون	ابيون
ترافق	برادش	اتار
ترافق	موسيق	اتجاه
ترافق	استراديول	اتنيل
ترافق	بروبين	اتين
ترافق	تشيرنكوف	اثر
ترافق	سطح	اثر
ترافق	فن	اثر
ترافق	هول	اثر
ترافق	عشر	اثنا
ترافق	غالينديس	اثنار
ترافق	الاا	اجا
ترافق	عمل	اجاز
ترافق	طاغوت	اجتنب
ترافق	قضا	اجتهاد
ترافق	مزنق	احتجاز
ترافق	كريم	احجار
ترافق	جانح	احداث
ترافق	مواد	اختبار
ترافق	لامارك	اخدرى
ترافق	صحف	اخراج
ترافق	فن	اخراج
ترافق	سبوت	اخرص
ترافق	سبيل	اخل
ترافق	سموم	ادمان
ترافق	سمندز	ادنغتن
ترافق	انسان	اذن
ترافق	بطين	اذين
ترافق	ارثوس	ارادوس
ترافق	بنا	اساس
ترافق	بعد	استشعار
ترافق	كهرساكن	اسقاط
ترافق	بيولوجي	اسلح

ترافق	ابيض	اسمنت
ترافق	امينت	اسمنت
ترافق	تجار	اسناد
ترافق	عرب	اسواق
ترافق	مال	اسواق
ترافق	شاق	اشغال
ترافق	حيا	اصل
ترافق	دين	اصلاح
ترافق	زراع	اصلاح
ترافق	نقد	اصلاح
ترافق	محاكم	اصول
ترافق	مسرح	اضا
ترافق	تاهيل	اعاد
ترافق	محاكم	اعاد
ترافق	مال	اعان
ترافق	عرض	اعتد
ترافق	مستند	اعتماد
ترافق	محل	اعذار
ترافق	سحيق	اعماق
ترافق	زراع	افات
ترافق	تعليم	اقتصاد
ترافق	رفا	اقتصاد
ترافق	زراع	اقتصاد
ترافق	سكان	اقتصاد
ترافق	طاق	اقتصاد
ترافق	عمل	اقتصاد
ترافق	غذا	اقتصاد
ترافق	قياس	اقتصاد
ترافق	معرف	اقتصاد
ترافق	افحاش	اقذاع
ترافق	طاير	اقلاع
ترافق	هبوط	اقلاع
ترافق	ارجاع	اكسد
ترافق	حشر	اكل
ترافق	لحوم	اكل

اعتمد في تقييم الأنطولوجيا العربية المولدة على البيانات Data driven من حيث عددها وصحتها، وذلك لكل من المفاهيم والمصطلحات والعلاقات. قيست صحة بيانات الانطولوجيا العربية المولدة بقياس الدقة Precision المستخدمة على نطاق واسع في تقييم نظم المعلومات وذلك بحساب حاصل قسمة العدد المتولد الموائم على العدد المتولد الكلي، وتم الاستعانة بخبراء من هيئة الموسوعة العربية لمعرفة العدد الموائم للمفاهيم والمصطلحات والعلاقات. يوضح الجدول رقم (3) نتائج التقييم. وقد كانت دقة المفاهيم 0.88 وهي عالية كونها مأخوذة من عناوين بحوث الموسوعة العربية والموضوعة من خبراء وباحثين عرب وعددها جيد مقارنة مع عدد مفاهيم

شبكة الكلمات العربية. كما أن عدد المصطلحات وعلاقاتها الهرمية مع المفاهيم والمستخرجة آلياً 12800 يعتبر جيداً بدقة 0.52. وقد كانت دقة علاقات الترافق بين المصطلحات عالية وهي 0.7. كما كانت دقة علاقات التشابه مقبولة وهي 0.4.

الجدول رقم (3) نتائج تقييم الأنطولوجيا العربية

	العدد الكلي	العدد الموائم	الدقة
المفاهيم	5778	5100	0.88
المصطلحات وعلاقاتها الهرمية مع المفاهيم	12800	6600	0.52
علاقات الترافق	3500	2450	0.7
علاقات التشابه	2500	1100	0.4

وتعتبر هذه النتائج جيدة لاستكمال وإغناء شبكة الكلمات العربية بطريقة آلية بإضافة مفاهيم ومصطلحات وعلاقات اعتماداً على مصدر معرفي عربي غني وغير مترجم للحصول على أنطولوجيا عربية تستخدم كنواة في التطبيقات الدلالية العربية. وقد تميزت هذه الطريقة عن المشاريع السابقة لتوليد أنطولوجيا عربية أنها آلية وليست يدوية وتعتمد على مصدر معرفي عربي وليس مترجم لتوسيع شبكة الكلمات العربية. كما أنها شاملة وليست متخصصة في مجال علمي معين.

الخاتمة والآفاق المستقبلية:

الهدف من هذا البحث هو إيجاد أنطولوجيا عربية. وقد وضع في بحثنا إطار لتوليد الأنطولوجيا العربية يتألف من ثلاثة مكوّنات أساسية هي: مكّون إدارة الأنطولوجيا من أجل تصفح وتعديل وصيانة الأنطولوجيا، ومكّون معالجة المصادر وتجهيزها وذلك من أجل استخراج خصائص الموسوعة العربية والعلاقات وتجهيز بيانات شبكة الكلمات العربية، ومكّون أداة خوارزمية تعلّم الأنطولوجية التي تطبق خوارزمية تعلم الأنطولوجيا بالتقابل بين خصائص الموسوعة العربية وعلاقاتها المستخرجة وبين بيانات شبكة الكلمات العربية المجهزة. وبناء على هذا الإطار تم إيجاد الأنطولوجيا العربية التي جمعت بين غنى شبكة الكلمات العربية وشموليتها وبين جودة مفاهيم الموسوعة العربية ومصطلحاتها وعلاقاتها كونها عربية أصلاً وغير مترجمة.

يمكن استخدام الأنطولوجيا العربية الناتجة في التطبيقات الدلالية على النصوص العربية مثل الوب الدلالي والتلخيص الآلي والترجمة الآلية وغيرها لتحسين جودة مخرجاتها باعتماد المفاهيم والعلاقات الدلالية العربية.

تعتبر الأنطولوجيا العربية المتولدة نواة لأنطولوجيا عربية شاملة يمكن تطويرها واستكمالها بإضافة مفاهيم وعلاقات جديدة إليها إما باستخدام تقنية التقابل مع خصائص مصادر معرفية عربية موثقة أخرى مثل المعاجم العربية، أو باستخدام تقنيات تعلم آلية أخرى. للحصول على أنطولوجيا عربية أكثر فاعلية وفعالية.

References

1. Aliane H., Alimazighi Z. Ahmed Cherif M. " Al –Khalil: The Arabic Linguistic Ontology Project" ,, Semantic web and Arabic Language Team, in Computer Science Department, University, Algiers, 2009.
2. AlBarhamtoshy H., Wajdi A., " .Designing and Implementing Arabic WordNet Semantic-Based". Language Engineering Conference.32− 23 .2009 .
3. Alkhalifa M., Rodríguez H., "Automatically Extending Named Entities coverage of Arabic WordNet using Wikipedia", International Journal on Information and Communication Technologies , vol 1, No. 1 2008
4. AlMaayah M., Sawalha M. Abushariah M." .Towards an automatic extraction of synonyms for Quranic Arabic WordNet ".International Journal of Speech Technology, Volume 19 :2016 ، .189−177
5. Alsudias L , Rayson P., "Developing an Arabic Infectious Disease Ontology to Include Non-Standard Terminology ", Proceedings of the 12th Conference on Language Resources and Evaluation ,4842–4850 , 2020.
6. Althobaiti M., Kruschwitz U., Poesio M. "Aranlp: A java-based library for the processing of arabic text" 2014 .
7. Asim M.-N., Wasim,M. Khan,M.U.G. ." .A survey of ontology learning techniques and applications ".Database.2018 ،
8. Bao J., Honavar V. .Collaborative Ontology Building with Wiki@nt A multi-agent based ontology building environment .Iowa State University.2004 ،
9. Belkredim Z., Meziane F." .DEAR-ONTO: A DErivational ARabic Ontology Based on Verbs ". International Journal of Computer Processing of Languages, 21 (3) 291-279 , 2008.
10. Benassi, Bergamaschi S. Vincini M. "Web Semantic Search with TUCUXI" the Twelfth Italian Symposium on Advanced Database Systems SEBD, 426–433, 2004..
11. Biemann C. B" .Ontology learning from text: A survey of methods ".LDV Forum 20 (2) 75-93, 2005
12. Black W. Farwell D. Elkateb S" .Arabic WordNet and the Challenges of Arabic ".Arabic NLP/MT Conference .London, U.K.2006 ،.
13. Carenini G., Cheung J. C. K." .Extractive vs. NLG-based abstractive summarization of evaluative text: the effect of corpus controversiality ".Fifth International Natural Language Generation Conference .PA, USA: Stroudsburg 33-41 , 2008
14. Cimiano P. C., "Ontology Learning and Population form Text: Algorithms, Evaluation and Applications", Springer, 2006.
15. Drumond L., Girardi R. "A Survey of Ontology Learning Procedures", WONTO, Vol. 427. 2008
16. Ehrig M., Sure Y. .Ontology Mapping − An Integrated Approach .Institute AIFB, University of Karlsruhe.2005 ،
17. Fawcett T. E .Feature discovery for problem solving systems .PhD Thesis ،Amherst, MA, USA: University of Massachusetts.1993 ،
18. Fellbaum C" .WordNet: an electronic lexical database ".Cambridge Mass: MIT Press.1998 ،
19. Fernando S., Stevenson M. .Mapping WordNet synsets to Wikipedia articles .University of Sheffield.2009 ،

20. Gómez-Pèrez, Asunción, Fernández-López, Mariano, Corcho, Oscar., "Ontological Engineering with examples from the areas of Knowledge Management" , e-Commerce and the Semantic Web, Springer Verlag, London , 2004.
21. Ishkewy H., Harb H., Farahat H, " .Azhary: An Arabic Lexical Ontology ".International Journal of Web & Semantic Technology (IJWesT).2014 ،
22. Jarrar M" .Building a Formal Arabic Ontology ".the Experts Meeting on Arabic Ontologies and Semantic Networks ..Alecso Arab League. Tunis.28−26 .2011 ،
23. Jarrar M., "The Arabic Ontology – An Arabic Wordnet with Ontologically Clean Content", Birzeit University, 2021
24. John P. M .Mapping WordNet Instances to Wikipedia .Ireland: Insight Centre for Data Analytics National University of Ireland Galway.2017 ،
25. Juzaddin A.., Saif O. N , , Mapping Arabic WordNet synsets to Wikipedia articles using monolingual and bilingual features .Cambridge University.2015 ،
26. Kanan R.A.S.G., Jaam J., Hailat E.," .Stop-word removal algorithm for arabic language ". International Conference on Information and Communication Technologies: From Theory to Applications. IEEE.2004 .
27. Mazari A. Ch., Aliane H. Alimazighi Z. .Automatic construction of ontology from Arabic texts ". ICWIT Conference on Web and Information Technologies, 2012.
28. Mohammed G.H., Al Zamil. Qasem Al-Radaideh." .Automatic extraction of ontological relations from Arabic text ".Journal of King Saud University– Computer and Information.2014 ،
29. Ponzetto S. P., Navigli R." .Knowledge-rich Word Sense Disambiguation rivaling supervised system48 ".th Annual Meeting of the Association for Computational Linguistics .،Sweden ، .1531−1522 .2010
30. Rodríguez H., Farwell D. Farreres J. Bertran M. Alkhalifa M. Martí M. A." .Arabic WordNet: Semi-automatic Extensions using Bayesian Inference6 ".th Conference on Language Resources and Evaluation LREC2008 .Marrak.2008 ،
31. Ruiz-Casado M., Alfonseca E. Castells P., "Automatic Assignment of Wikipedia Encyclopedic Entries to WordNet Synsets. Advances in Web Intelligence ".Third International Atlantic Web Intelligence Conference.2005 .
32. Soor S .ONTOLOGY MAPPING .India: 8th Mile, Mysore Road, RVCE Post.2015 ،
33. Smith B .Ontology .Blackwell: Malden: Floridi (Ed.), The Blackwell Guide to Philosophy of Computing and Information.2004
34. Suchanek F. M., Kasneci G. Weikum G." .YAGO: A Large Ontology from Wikipedia and WordNet ".Journal of Web Semantics.2008 ،.
35. Szymanski J.., Korytkowski R. Krawczyk H. "Collaborative Approach to WordNet and Wikipedia Integration ".The Second International Conference on Advanced Collaborative Networks, Systems and Applications.2012 .
36. Van Rijsbergen" .File Organization In Library Automation And Information Retrieval ".Journal of Documentation, Vol. 32 No. 4.317−294 :1976 ،،