

## أثر استخدام تقنيات معالجة اللغة الطبيعية للويب الدلالي في صناعة المعاجم المتخصصة الآلية: دراسة تطبيقية في المصطلحات التقنية.

### The Impact of Using Natural Language Processing Techniques for The Semantic Web Development of Specialized Automatic Dictionaries: An Applied Study in Technical Terminology

يسمينه قالي\* ، مخبر التداولية واستراتيجيات الخطاب، جامعة محمد لمين دباغين - سطيف2،

الجزائر، [y.gali@univ-setif2.dz](mailto:y.gali@univ-setif2.dz)

خالد هدنة\* ، مخبر التداولية واستراتيجيات الخطاب، جامعة محمد لمين دباغين - سطيف2،

الجزائر، [k.hedna@univ-setif2.dz](mailto:k.hedna@univ-setif2.dz)

تاريخ قبول المقال: 17-03-2025

تاريخ إرسال المقال: 09-01-2025

**المخلص:** تستكشف هذه الدراسة طرق تمثيل المعرفة المتخصصة من خلال توظيف تقنيات معالجة اللغة الطبيعية والويب الدلالي. وتركز على تطوير المعاجم الآلية المتخصصة من خلال نمذجة العلاقات الدلالية بين المصطلحات ومفاهيمها، ما يساهم في تحسين استخراج المصطلحات وتحليل النصوص المتخصصة. كما تسعى إلى توظيف المعلومات التوثيقية المهيكلية لبناء قاعدة مصطلحية تقنية متخصصة، ترتبط بقواعد بيانات معرفية خارجية، والهدف من ذلك تسهيل الوصول إلى الموارد المصطلحية، وتحسين أنظمة ترجمة الويب باستخدام البيانات المترابطة.

**الكلمات المفتاحية:** معالجة اللغة الطبيعية، الويب الدلالي، المعاجم الآلية، نمذجة العلاقات الدلالية.

**Abstract:** This study explores methods for representing specialized knowledge by employing Natural Language Processing (NLP) and semantic web techniques. It focuses on developing specialized computational lexicons by modeling the semantic relationships between terms and their corresponding concepts, contributing to more efficient term extraction and specialized text analysis. the study also aims to leverage structured linguistic metadata to build a specialized terminological database interconnected with external knowledge bases. the ultimate goal is to facilitate access to specialized lexical resources and enhance machine translation systems using linked data.

**Key words:** natural language processing, Semantic Web, computational lexicons, modeling semantic relationships.

**مقدمة:**

إنّ فهم وتفسير مصطلحات العلوم التقنية يعدّ أمراً حيوياً في مجالات مثل الهندسة والرعاية الصحية والتقنية، وهو ما يستدعي توحيد الجهود العربية في مجال العمل المصطلحي والترجمي، وكذلك لأجل تسهيل التواصل العلمي بين الباحثين والمتخصصين في ذات المجال، وتتمثل خصوصية المصطلحات التقنية في تعقيد أنظمتها المفاهيمية، لذا فإن توفير معاجم متخصصة تعكس هذا التعقيد، يعدّ خطوة أساسية لضمان تبادل المعلومات بسلاسة وتجنب الإرباك الذي يمكن أن يخلقه التعدد المصطلحي للمفهوم الواحد. وقد اكتسب الارتباط بين تقنيات الويب الدلالية والمعاجم المتخصصة أهمية كبيرة، لأنه يستخدم معلومات الويب ويجعلها قابلة للتمثيل والمعالجة الآلية، ويستعين بعدة مجالات مثل الأنطولوجيا، والمنطق، والاستدلال، ما يسمح بتكليف المصطلحات وتمثيلها بطريقة منظمة ومتراصة دلالياً. ويهدف الويب الدلالي إلى تمكين الآلة من فهم محتوى وسياق موارد الويب بشكل أفضل، مما يسهل معالجة المعلومات ومشاركتها، كما تتيح التقنيات الدلالية تشفير البيانات النصية، وترتكز على تمثيل المعاني والعلاقات بين الكلمات والجمل بدلاً من التركيز على التركيب النصي فقط، وهذا يساعد في خفض نسبة اللبس عند الترجمة أو تحليل النصوص، ومن ناحية أخرى تعد معالجة اللغات الطبيعية فرعاً من فروع الذكاء الاصطناعي، وقد مهدت هذه التقنيات السبيل لإحراز تقدم عميق في تمثيل واستخدام المصطلحات المتخصصة، وإعادة هيكلة واستغلال القواميس وبناء المعاجم المتخصصة.

وعلى ضوء ما سبق يطرح البحث مجموعة من التساؤلات منها: كيف تعمل تقنيات الويب الدلالية على تيسير التعرف على المصطلحات المتخصصة؟ وكيف يمكن أن تسهم تقنيات الويب في تخطي الحواجز الثقافية واللغوية في ترجمة مجالات العلوم المختلفة؟

وما دور هذه التقنيات في مجال أتمتة مهام مثل: استخراج المصطلحات، التغلب على مشكلة الغموض المصطلحي؟ وكيف تتم عملية إثراء مصطلحات العلوم التقنية بالمعلومات الإضافية المترابطة بالسياقات الاستعمالية في النصوص المتخصصة؟

انطلاقاً مما سبق؛ تقترح الدراسة إطاراً دلالياً يدمج بين تقنيات الويب الدلالي مع نماذج معالجة اللغة الطبيعية الخاصة بالمجال لتطوير المعاجم الآلية، نقوم خلالها بمعالجة مدونة تقنية كنموذج رئيسي مع تحليل النتائج، ويبرز هذا المجال كحقل يتسم بالتوسع المستمر، ما يستدعي استخدام أساليب ديناميكية تعتمد على الدلالات في تمثيل المصطلحات وتطوير المعاجم المتخصصة بصورة أكثر دقة وشمولاً.

**المبحث الأول: معالجة اللغة الطبيعية للويب الدلالي: مفاهيم أساسية****المطلب الأول: الويب الدلالي: مفهومه وتقنياته.**

يسلّط هذا الجزء الضوء على مفهوم الويب الدلالي وتقنياته، وآثارها على تعزيز تمثيل المعرفة، وذلك في النصوص المتخصصة، ويمكننا تعريف الويب الدلالي بقول ديانا ماينارد (Diana Maynard) بأنه: «إضافة الدلالات، أو المعاني، إلى البيانات الموجودة على شبكة الإنترنت، بحيث يمكن معالجة صفحات الويب والتعامل معها من قبل الآلة بسهولة كبرى، وأحد المظاهر الرئيسة لهذا المفهوم تتمثل في وصف الموارد باستخدام مُعرّفات فريدة، تسمى مُعرّفات الموارد الموحدة. (URIs) يمكن أن تكون الموارد كيانات أو علاقات تصف كيفية ارتباط الكيانات بعضها ببعض»<sup>1</sup>. ويعتبر تيم بيرنرز لي (Tim Berners-Lee) واحداً من أهم مطوري شبكات الويب الدلالي، حيث قام بوضع إطار عملي أواخر عام 2004، والتي يمكن وصفها بأنها نقلة نوعية في تمثيل المعلومات وتبادلها، و: «يُعدّ امتداداً للويب - الحالي<sup>2</sup> - (الويب الثابت (Static Web) )، حيث يتم إعطاء المعلومات معنى محدداً بوضوح، مما يتيح لأجهزة الكمبيوتر والأشخاص العمل بشكل تعاوني وبشكل أفضل»<sup>3</sup>.

حسب ما سبق، فإن المعالجة الدلالية للويب تشمل ما يأتي:

- إضافة الدلالات والمعاني: وتتضمن توفير معلومات إضافية تقدم وصفا للمفردات والمفاهيم وسياقاتها، ويصطلح عليها بالبيانات الوصفية.
- تصنيف البيانات: ويشير إلى تنظيم وتصنيف البيانات على الويب بحيث يمكن تجميعها في فئات أو مواضيع مشابهة، وتشمل البيانات المفاهيم والعلاقات والسمات وما إلى ذلك<sup>4</sup>.
- الوصف باستخدام مُعرّفات فريدة (URIs): وهي عبارة عن مُعرّفات أو روابط تُستخدم لتمثيل كائن أو مفهوم معين على الويب. مثل عناوين (URL) يتم استخدامه في سياق البيانات المترابطة.

<sup>1</sup>. ماينارد ديانا وآخرون، معالجة اللغات الطبيعية للويب الدلالي، تر: خالد بن عبد الرحمن الميمان، ط1، مركز الملك عبد الله بن عبد العزيز الدولي لخدمة اللغة العربية، الرياض، 2019م، ص: 23.

<sup>2</sup>. القصد من ذلك أنه قبل ظهور الويب الدلالي، كان بحث الويب يتم استناداً إلى مصطلحات البحث (keywords)، والعبارات الموجودة في صفحات الويب. عندما يقوم مستخدم بإدخال مصطلح بحثي معين في محرك بحث، كان يتم فحص صفحات الويب للبحث عن تلك المصطلحات ومطابقتها مع المحتوى الموجود في الصفحات. هذا النموذج البسيط للبحث كان يعتمد على مفهوم التتابع النصي فقط، ولم يكن فهم حقيقي للمعنى الحقيقي للكلمات أو العلاقات بينها. لذلك، لم يكن يوفر نتائج دقيقة ومعتمدة.

<sup>3</sup>. Berners-Lee, Tim, and al. "The Semantic Web". Scientific American, Scientific American Inc., Vol. 284, No. 5, 2001.

<sup>4</sup>. Bellandi Andrea. "Le Risorse Linguistiche nell'era del Web Semantico". AIDAinformazioni, Cacucci Editore, Vol. 40, no. 1-2, 2022, p. 33.

- **الموارد (كيانات أو علاقات):** وتشمل مجموعة متنوعة من البيانات والأدوات التي تستخدم لتمثيل اللغة وفهمها بدقة على الويب. مما يدعم تطبيقات: كمعالجة اللغة الطبيعية ومحركات البحث الذكية، يمكن أن تكون هذه الموارد: كيانات، مثل «بارك أوباما»، أو مفاهيم مثل «سياسي» أو علاقات تصف كيفية ارتباط الكيانات بعضها ببعض، مثل «زوجة»<sup>1</sup>.

ويتم تحقيق ذلك من خلال استخدام تقنيات الويب الدلالي، مثل RDF و OWL، بناءً على مجموعات فرعية من المنطق الوصفي<sup>2</sup>. ويتم جعل البيانات قابلة للوصول عبر نموذج البيانات المترابطة، حيث يتم تزويد كل كيان من مجموع البيانات (المفاهيم والعلاقات والسمات وما إلى ذلك) بمعرفات موحدة (IRI) ما يتيح الوصول إليها عبر الويب باستخدام بروتوكول نقل النص التشعبي (HTTP)<sup>3</sup>. يمكن التفصيل في ذلك كما يأتي:

**أولاً: RDF (Resource Description Framework):** هو إطار عمل لوصف المعلومات على الويب باستخدام العلاقات بين الموارد، بما في ذلك المستندات والأشخاص والأشياء والمفاهيم. يتمحور المفهوم الأساسي لـ RDF حول تمثيل البيانات باستخدام ثلاثيات؛ حيث يعبر بيان RDF عن علاقة بين موردين. يمثل الموضوع والكائن المصدرين اللذين هما ذات الصلة؛ يمثل المسند طبيعة علاقة. تتم صياغة العلاقة بطريقة موجهة من الموضوع إلى الكائن ويسمى في RDF خاصية<sup>4</sup>:

**<subject> <predicate> <object>**

لنأخذ الجملة التالية " يمكن للإطارات المتطورة تقليل الاحتكاك بالطرق وتقليل مقاومة الدوران، وبالتالي توفر استهلاك الوقود"<sup>5</sup> مثالا.

لتمثيل العلاقات بين **المواضيع (Subjects)** و**المسند (Predicates)** و**الكائنات (Objects)**، تحول العبارة السابقة إلى ثلاثيات RDF باستخدام مفهوم الموضوع (subject) والمسند (predicate) والكائن (object) على النحو التالي:

1- ثلاثي RDF الأول: الإطارات المتطورة تقلل الاحتكاك بالطرق

<sup>1</sup>. ينظر: ماينارد ديانا وآخرون، معالجة اللغات الطبيعية للويب الدلالي، ص: 23.

<sup>2</sup>. Baader, Franz, and al. THE DESCRIPTION LOGIC HANDBOOK: Theory, implementation, and applications. Cambridge: Cambridge university press. 2017, p. 8.

<sup>3</sup>. Bellandi Andrea. "Le Risorse Linguistiche nell'era del Web Semantico". P. 33. ينظر.

<sup>4</sup>. Schreiber, Guus, and Yves Raimond. "RDF 1.1 Primer". W3C, 2014, Available at: <https://www.w3.org/TR/rdf11-primer/>. Accessed 11 August 2023.

<sup>5</sup>. يسري مصطفى وآخرون، الطاقة الشمسية: النظرية والتطبيق، ط1، مؤسسة الأمة العربية، القاهرة، 2020، ص: 36.

**<الإطارات المتطورة> <تقليل> <الاحتكاك بالطرق>**

حيث: - مفهوم الموضوع (Subject) : "الإطارات المتطورة". - مفهوم المسند (Predicate) : "تقليل".  
- مفهوم الكائن (Object) : "الاحتكاك بالطرق".

2- ثلاثي RDF الثاني: الإطارات المتطورة تقليل مقاومة الدوران.

**<الإطارات المتطورة> <تقليل> <مقاومة الدوران>**

حيث: - مفهوم الموضوع (Subject) : "الإطارات المتطورة". - مفهوم المسند (Predicate) : "تقليل".  
- مفهوم الكائن (Object) : "مقاومة الدوران".

3- ثلاثي RDF الثالث: الإطارات المتطورة توفير استهلاك الوقود.

**<الإطارات المتطورة> <توفير> <استهلاك الوقود>**

حيث: - مفهوم الموضوع (Subject) : "الإطارات المتطورة". - مفهوم المسند (Predicate) : "توفير".  
- مفهوم الكائن (Object) : "استهلاك الوقود".

تستخدم معرفات الموارد الدولية (IRIs) لتمثيل الموارد على الويب بطريقة عالمية ومتعددة اللغات. وهي مشابهة لمعرفة الموارد الموحدة (URIs)، والفارق الرئيسي بينهما هو أن (URIs) القدرة على تمثيل الأحرف والرموز من مختلف اللغات والمجموعات الأخرى في IRIs<sup>1</sup>. وبالعودة إلى النموذج الذي أوردناه سابقاً، لاحظ أنه يمكن استخدام IRIs كمعرفات للموارد والعلاقات في RDF، على سبيل المثال؛ "http://example.org/ontology/Frame" هو: IRI يُستخدم لتمثيل نوع المورد "إطار Frame" في الأنطولوجيا المستخدمة، والموارد الأخرى مثل "الإطارات المتطورة" و"الاحتكاك بالطرق" وغيرها هي أيضاً معرفة باستخدام IRIs.

**ثانياً: OWL (Web Ontology Language):** هي لغة تمكن من تحديد وتمثيل العلاقات والتصنيفات بين المفاهيم على الويب. تُستخدم OWL لبناء الأنطولوجيا، وهي هيكل تصنيفي يصف العلاقات بين المفاهيم والكيفية التي يمكن تصنيفها بها<sup>2</sup>، وتستند في ذلك على RDF، من خلال توصيف المفاهيم والكائنات والعلاقات بينها، باستخدام الصفوف (Classes) والخصائص<sup>3</sup> (Properties).

<sup>1</sup>. Schreiber Guus, and Yves Raimond. "RDF 1.1 Primer", p: 33. ينظر:.

<sup>2</sup>. W3C. "OWL 2 Web Ontology Language: Overview". 2012, Available at: <https://www.w3.org/TR/owl2-overview/#ref-owl-2-specification>. Accessed 11 August 2023.

<sup>3</sup>. يعد OWL2 ملحقات RDF Schema، بالمعنى الذي يتبناه OWL2 لمعاني RDFS للفصول والخصائص (مثل rdfs: subClassOf، Class، وما إلى ذلك) ويضيف عناصر لغوية لدعم التعبير الأغنى المطلوب. يمكن أن يكون هذا النهج متسقاً مع الهيكل المتدرج للويب الدلالي.

والقيام بالاستدلال الآلي لفهم البيانات اللغوية واستنتاج العلاقات الفعلية والضمنية داخل البيانات اللغوية. مثلاً؛ يمكننا توصيف مفهوم "الألواح الشمسية" باستخدام لغة xml كالتالي:

1- تعريف الفئة (Class): يمكن تعريف فئة تحمل اسم "الألواح الشمسية" لتمثيل "الألواح الشمسية" ككائنات في المفهوم.

الشكل 1: نموذج لكيفية تعريف الفئة (class) باستخدام الـ OWL.

```
<owl:Class rdf:about="#SolarPanel">
  <rdfs:label>الألواح الشمسية</rdfs:label>
</owl:Class>
```

2- تعريف الخصائص (Properties): يمكن تعريف الخصائص التي تصف "الألواح الشمسية"، مثل: "القدرة الكهربائية" و"المواد المستخدمة" وما إلى ذلك.

الشكل 2: نموذج لكيفية تعريف الخصائص (properties) باستخدام الـ OWL.

```
<owl:ObjectProperty rdf:about="#hasPowerCapacity">
  <rdfs:label>لها قدرة كهربائية</rdfs:label>
  <rdfs:domain rdf:resource="#SolarPanel"/>
  <rdfs:range
    rdf:resource="http://www.w3.org/2001/XMLSchema#double"/>
</owl:ObjectProperty>
<owl:ObjectProperty rdf:about="#madeOfMaterial">
  <rdfs:label>مصنوعة من مادة</rdfs:label>
  <rdfs:domain rdf:resource="#SolarPanel"/>
  <rdfs:range
    rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:ObjectProperty>
<owl:ObjectProperty rdf:about="#locatedIn">
  <rdfs:label>موجودة في</rdfs:label>
  <rdfs:domain rdf:resource="#SolarPanel"/>
  <rdfs:range
    rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:ObjectProperty>
```

**ملاحظة:** تستخدم لغة الـ RDF لتمثيل العلاقات البسيطة بين المفاهيم، فيما تستخدم لغة الـ OWL لتوضيح العلاقات المعقدة بين المفاهيم، وتقديم بيانات تفصيلية أكثر لفهم العلاقات المنطقية التي تربط بين هذه المفاهيم وتفسيرها، خصوصاً تلك التي تنتمي لمجال واحد أو لمجالات ذات صلة ببعضها البعض.

**ثالثاً: SPARQL (SPARQL PROTOCOL and QUERY LANGUAGE):** هي

عبارة عن لغة استعلام تسمح بالبحث داخل قواعد البيانات، واسترجاع البيانات المخزنة ومعالجتها من

مخازن بيانات الـ RDF. يمكنك استخدام SPARQL للبحث عن بيانات محددة بناءً على العلاقات والمفاهيم<sup>1</sup>.

رابعاً: البيانات المترابطة (Linked Data): تقنية Linked Data تسعى إلى توصيل وربط مصادر المعلومات على الويب بشكل موحد باستخدام RDF وعناوين الـ URI. هذا يمكن من إنشاء شبكة متصلة من المعلومات تتيح الوصول إلى البيانات من مصادر متعددة<sup>2</sup>.

المطلب الثاني: معالجة اللغة الطبيعية (NLP): سنتناول في هذا الجزء مفهوم معالجة اللغات الطبيعية وأهميتها وتطبيقاتها العملية في تحليل النصوص وفهمها.

### أولاً: تعريف المعالجة الآلية للغة الطبيعية وأهميتها:

تعرف معالجة اللغة الطبيعية<sup>3</sup> بأنها: «ممارسة متطورة في علوم الكمبيوتر والذكاء الاصطناعي (AI) وتعنى بمعالجة اللغات الطبيعية مثل الإنجليزية أو الماندرين<sup>4</sup>، إذ تشمل هذه المعالجة عادة ترجمة اللغة الطبيعية إلى بيانات (أرقام) يمكن للكمبيوتر استخدامها للتعرف على العالم. يُستخدم هذا الفهم للعالم أحياناً لإنتاج نصوص باللغة الطبيعية تعكس هذا الفهم<sup>5</sup>. من خلال هذا التعريف؛ نجد أنّ معالجة اللغة الطبيعية تعنى أساساً بفهم اللغات الانسانية ومحاولة تطويعها وتحويلها إلى نموذج يسهل فهمه من قبل برامج الحاسوب، ثم تستخدم نتائجها وتطبيقاتها في تطوير أنظمة التدقيق اللغوي، واستخلاص واسترجاع المعلومات آلياً، والترجمة الآلية، وفهم النصوص وتصنيفها، وبناء بنوك وقواعد البيانات المعجمية، والتعرف على الكلام، بالإضافة إلى روبوتات الدردشة، وبناء بنوك وقواعد البيانات المعجمية... وغيرها، كما تهدف لتعزيز "الوصول إلى نظم المعلومات في مجالات المعرفة الإنسانية"<sup>6</sup>.

<sup>1</sup> Harris, Steve, and Seaborne Andy, "SPARQL 1.1 Query Language", W3C, 2013, Available at: <https://www.w3.org/TR/sparql11-query/>, Accessed 11 August 2023.

<sup>2</sup> Tim Berners-Lee, "linked data", 2009, Available at: <https://www.w3.org/DesignIssues/LinkedData>, Accessed 11 August 2023.

<sup>3</sup> وتُعرف أيضاً باسم اللغويات الحاسوبية (CL) أو هندسة اللغات الطبيعية (NLE)، حسب: ماينارد ديانا وآخرون، معالجة اللغات الطبيعية للويب الدلالي، ص: 23.

<sup>4</sup> اللغة الصينية الرسمية.

<sup>5</sup> Hobson Lane and al., Natural Language Processing in Action, 2nd edition, Manning Publications, New York, 2021, p: 12.

<sup>6</sup> ينظر: جاد الكريم عبد الله، عبقرية العربية والذكاء الاصطناعي - آفاق وطموحات -، ط1، دار النابغة للنشر والتوزيع، طنطا، 2020، ص: 288.

**ثانياً: تقنيات معالجة اللغة الطبيعية:** وتشمل مجموعة واسعة من التقنيات مثل: «تحليل النص وتقسيمه إلى: جمل وكلمات، وإضافة الحواشي والشروحات الدلالية واستخراج الآراء والتوجيهات»<sup>1</sup>. ويمكننا إضافة: معالجة النص المسبقة، تحليل المشاعر، استخراج المعلومات، تصنيف الموضوعات.. إلخ..

**المطلب الثالث: منهجية البحث:** لرصد أثر استخدام تقنيات معالجة اللغة الطبيعية للويب الدلالي في صناعة المعاجم المتخصصة الآلية، ارتأينا تحليل عينة نصية مقتبسة من دليل استخدام طابعة CANON LBP6030 -النسخة العربية-<sup>2</sup>، ولكن لطول المدونة اخترنا ثلاث فقرات كأمثلة للتوضيح مع إجراء بعض التعديلات عليها، حيث أن أدلة الاستخدام تستعمل لغة تقنية في التحرير بسبب تناولها لطريقة تركيب واستعمال المنتجات في مجالات مختلفة. وتشمل الدراسة التطبيقية عدة مراحل هي: التحليل النصي، استخراج المصطلحات، إثراء المصطلحات بالمعلومات الوصفية، ثم وصف وتحليل العلاقات بين المصطلحات، وقد تم اعتماد لغة بايثون في معالجة النصوص، إلى جانب الاستعانة بمكتبات برمجية مختلفة حسب المهام المطلوبة في كل مرة.

**نصوص المدونة:** تمثلت هذه النصوص في العينات الآتية:

- التَّنْبِيْثُ: لَا تَقُمْ بِتَنْبِيْثِ الْجِهَازِ فِي مَكَانٍ قَدْ يُؤَدِّي إِلَى حُدُوْثِ حَرِيْقٍ أَوْ صَدْمَةٍ كَهَرَبَائِيَّةٍ.
- إِمْدَادُ الطَّاقَةِ: هَذَا الْجِهَازُ مُتَوَافِقٌ مَعَ جُهْدٍ كَهَرَبَائِيٍّ مِنْ 220 إِلَى 240 فُولْتٍ وَتَرَدِّدٍ كَهَرَبَائِيٍّ 50/60 هِرْتِزَ.
- يَجِبُ تَوْصِيْلُ الْجِهَازِ بِمِقْبَسٍ مَأْخُذٍ مُرَوِّدٍ بِوَسْلَةٍ تَأْرِیْضُ بِاسْتِخْدَامِ سِلْكِ النَّيَّارِ الْكَهَرَبَائِيِّ الْمُرَوِّدِ.

**المبحث الثاني: الدراسة التطبيقية:** يشتمل المبحث على ثلاث مراحل أساسية: مرحلة تحليل النصوص وتحديد أجزاء الكلام، مرحلة استخراج المصطلحات، والمرحلة الأخيرة تمثيل المصطلحات والعلاقات دلالية.

<sup>1</sup>. ينظر: ماينارد ديانا وآخرون، معالجة اللغات الطبيعية للويب الدلالي، ص: 23.

<sup>2</sup>. تم نشر الملف بتنسيق (search.json)، ويمكن تحميل دليل استخدام الطابعة من موقع الشركة على الويب، كما يمكن الاطلاع على محتوى الوثيقة بالتنسيق النصي في البوابة الالكترونية لدليل استخدام الطابعة، مدخل "تعليمات أمان مهمة":

<https://oip.manual.canon/USRMA-3694-zz-SS-ar/contents/01020000.html>

## المطلب الأول: مرحلة تحليل النصوص، وتحديد أجزاء الكلام:

وتعرف أيضا بتقسيم النص (Tokenization)، والهدف من معالجة النص هو "تحليله إلى جمل<sup>1</sup> ثم كلمات، والتخلص من علامات الترقيم الملتصقة بالكلمات"<sup>2</sup> مثل الفواصل والنقاط وعلامات التنصيص.. وغيرها، واستخراج المداخل المعجمية، ثم تحليلها صرفيا إلى مادتها المعجمية (Lemma)<sup>3</sup>، وتحديد جذر الكلمة (Root)، وتحديد قسمها الكلامي (part of speech)<sup>4</sup>. فالمصطلحات "تتنظم في مجموعة محدودة نسبيا من البنى الصرفية والمعجمية (مصطلح مفرد، مصطلح مركب)"<sup>5</sup>، وأما أقسام الكلام فتساعدنا في التعرف على التراكيب المصطلحية<sup>6</sup> المختلفة التي غالبا ما ترد كمركبات اسمية إذا استثنينا التعابير الاصطلاحية. باستخدام لغة Python وأداة (SinaTools)<sup>7</sup> للتحليل الصرفي (ALMA)، عند تشغيل كود التحليل حصلنا على النتائج موضحة في الشكل الآتي:

الشكل 3: نموذج لمخرجات التحليل الصرفي للمدونة النصية باستخدام مكتبي nltk و spaCy.

```
# تحليل الجملة الأولى
data_1 = [
  {'token': 'الْبَيْتُ', 'lemma': 'بَيْت', 'root': 'ب', 'pos': 'اسم', 'frequency': 511},
  {'token': ':', 'lemma': ':', 'root': ':', 'pos': 'علامة ترقيم', 'frequency': 0},
  {'token': 'لا', 'lemma': 'لا', 'root': 'لا', 'pos': 'أداة نفي', 'frequency': 11215106},
  {'token': 'تَقْمُ', 'lemma': 'تَقْم', 'root': 'ق', 'pos': 'فعل', 'frequency': 1707567},
  {'token': 'يَنْتَبِهُمُ', 'lemma': 'يَنْتَبِهُ', 'root': 'ن', 'pos': 'اسم', 'frequency': 5116},
  {'token': 'جَز', 'lemma': 'جَز', 'root': 'ز', 'pos': 'اسم', 'frequency': 69625},
  {'token': 'نبي', 'lemma': 'نبي', 'root': 'ن', 'pos': 'حرف جر', 'frequency': 63929790},
  {'token': 'مَكَان', 'lemma': 'مَكَان', 'root': 'ك', 'pos': 'اسم', 'frequency': 852272},
  {'token': 'أداة قلم', 'lemma': 'أداة قلم', 'root': 'أ', 'pos': 'أداة قلم', 'frequency': 4675944}
]
```

<sup>1</sup>. يتم في معظم الحالات تحليل النصوص إلى كلمات مباشرة دون المرور على التقسيم الجملي، ويمكن أن يتم الفصل بينها في النتائج عند نهاية كل جملة أو فقرة، لذا يمكن تصحيح ذلك بالقول إنه يتم تحليل النص إلى وحدات أصغر، مثل: الكلمات والرموز.

<sup>2</sup>. ينظر: الفيفي عبد الله بن يحيى، مدخل إلى اللسانيات الحاسوبية، ط1، مركز الملك عبد الله الدولي لخدمة اللغة العربية، الرياض، 2017م، ص: 139.

<sup>3</sup>. ينظر: اللتيني يحيى أحمد، أسس صناعة المعجم المحسوب، ط1، كنوز المعرفة، الأردن، 2019م، ص: 149.

<sup>4</sup>. ينظر: اللتيني أحمد يحيى، أسس صناعة المعجم المحسوب، ص: 199.

<sup>5</sup>. أكماضان محمد، "لمحة عامة عن الأسس النظرية والمنهجية للعمل المصطلحي"، دائرة الترجمة العربية، نيويورك، 2015م، ص: 06.

<sup>6</sup>. ينظر: العنزي محمد بن نافع، المصاحبة اللفظية في العربية المعاصرة وأثرها في تغير الدلالة، مجلة الدراسات اللغوية، مركز الملك فيصل للبحوث والدراسات الإسلامية، م15-ع01، 2012م، ص: 20.

<sup>7</sup>. أدوات SinaTools : هي مجموعة أدوات مفتوحة المصدر لمعالجة اللغة الطبيعية تدعم اللغة العربية؛ تتضمن أدوات ل: التوسيم الصرفي، إزالة الغموض عن معاني الكلمات، التعرف على الكيانات المسماة، استخراج المرادفات، العلاقات الدلالية... إلخ. وقد تم تطويرها بواسطة فريق SinaLab جامعة بيرزيت- فلسطين. تتوفر SinaTools من خلال واجهات برمجة تطبيقات Python

وسطور الأوامر والمختبرات التعاونية والعروض التوضيحية عبر الإنترنت، ينظر: تاريخ التصفح 03 سبتمبر 2023  
<https://sina.birzeit.edu/sinatools/documentation/index.html>

**التعليق والتحليل:** يعرض الشكل (3) أعلاه المداخل المعجمية التي نتجت عن تحليل نصوص المدونة، مع تحليلها الصرفي، والأقسام النحوية التي تنتمي إليها (ينظر الجدول (1))، حيث تعتمد الأداة على قاموس الحلول المورفولوجية<sup>1</sup>، ويتم ترتيبها وفق نسبة تكرارها داخل القاموس<sup>2</sup>، أي أن الأداة تعتمد في عمليتي التحليل والتصنيف على مبدأ الشيوخ.

**Lemma :** لاحظ كلمة "إمداد"، لم يتم اختزالها كما يفترض، لكن تم إيرادها بصيغة جمع المؤنث السالم "إمدادات"، وفق لمبدأ الشيوخ الذي يستند إليه عمل الأداة.

**POS :** كلمة "كهربائية" تم تصنيفها "اسما" بدل "صفة"، كما وردت في المدونة النصية في سياقاتها المختلفة، إلا أننا إذا اطلعنا عليها من خلال معرفها الرقمي داخل القاموس، سنجد من خلال اشتقاقاتها : كهرباء < اسم، كهربائي < صفة، كهربائية < اسم<sup>3</sup>، وغالبا فإن هذه صورها الشائعة التي وردت بها داخل المدونات النصية التي تم الاعتماد عليها لبناء أداة التحليل (ALMA). وينطبق الأمر ذاته على مفردات: حريق، مأخذ، سلك < صنف أسماء بدل صفات، و <متوافق> صنف الكلمة صفة بدل اسم.

**الجدول 1: نموذج من مخرجات التحليل النصي للمدونة مع توضيح المكونات اللغوية المستخرجة.**

|                                                                                  |                                                                                                                                                                                                                                                                                                                                                      |
|----------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| [<br>'token':<br>'التثبيت',<br>'lemma':<br>'تثبيت',<br>'lemma_id':<br>202008767, | <b>Tokens :</b> ترمز للكلمة كما وردت في النص الذي تم تحليله                                                                                                                                                                                                                                                                                          |
|                                                                                  | <b>Lemma :</b> الشكل الأساسي للكلمة، مترجم إلى مصطلح المادة المعجمية <sup>4</sup> .                                                                                                                                                                                                                                                                  |
|                                                                                  | <b>Lemma_id :</b> المعرف الرقمي للمادة المعجمية في المعجم الذي ترتبط به أداة التحليل، وهنا ترتبط بمعجم قبس الإلكتروني، يسمح بالوصول المباشر إلى المدخل المعجمي (كلمة "تثبيت" مثالا) ، في قاعدة البيانات المعجمية، لاحظ رابط المفردة :<br><a href="https://sina.birzeit.edu/qabas/lemma/202008767">https://sina.birzeit.edu/qabas/lemma/202008767</a> |

<sup>1</sup>. وهي السمات الصرفية والنحوية للكلمة: كالجنس، والعدد، والوزن، الحالة الإعرابية، والزمن... وغيرها، والتي تستند إليها أدوات التحليل اللغوي.

<sup>2</sup>. ينظر: Jarrar Mustafa, and al. "ALMA: Fast Lemmatizer and POS Tagger for Arabic". *Procedia Computer Science*, 244, 2024, p: 1.

<sup>3</sup>. ينظر: مدخل كلمة "كهربائي"، معجم قبس، تاريخ التصفح: 03 سبتمبر 2025:

<https://sina.birzeit.edu/qabas/lemma/202008767>

<sup>4</sup>. ينظر: اللتيني يحيى أحمد، أسس صناعة المعجم المحوسب، ص: 149.

|                                                |                                                                                                                                                                                                                          |
|------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>'root': ' ث</p> <p>'pos':</p> <p>'اسم',</p> | <p><b>Root</b> : هو الجذر الصرفي للمفردة، ويجب التفريق بينه وبين الجذع Stem والليما؛ حيث أن الجذع هو الكلمة التي تم حذف سوابقها ولواحقها باستخدام التحليل الصرفي، مع العلم أنه قد ينتج عن ذلك جذوع غير متداولة لغويا</p> |
| <p>'frequency':</p>                            | <p><b>Pos</b> : القسم النحوي الذي تنتمي إليه المفردة في أقسام الكلام (اسم، فعل، حرف، صفة... إلخ).</p>                                                                                                                    |
| <p>51169</p> <p>{,}</p>                        | <p><b>Frequency</b> : التكرار أو التردد، أي : عدد المرات التي تكررت فيها الكلمة في سياقات الاستعمال في المدونة التي تم الاستناد إليها أثناء تدريب الأداة.</p>                                                            |

**الملاحظ** أيضا أنّ الأداة لم تتعرف على كلمات "قولت" و"هيرتز" كمفردات غير عربية، فهي غير خاضعة لقواعد العربية، لأنها وحدات قياس. ويمكن تقييم دقة عمل الأداة في التحليل كما يلي: تحليل الكلمات إلى شكلها القاعدي: 93.47%، استخراج الجذور: 95.65%، تصنيف أجزاء الكلام: 80.43%، مع التوضيح أن هذه النسب حسبناها وفق المدونة التي عملنا عليها دون الأخذ في الاعتبار مبدأ الشيوخ الذي يستند إليه عمل الأداة.

#### المطلب الثاني: مرحلة استخراج المصطلحات:

في هذه المرحلة سنقوم بالاعتماد على مخرجات المرحلة السابقة التي تربط المفردات بنوعها في أقسام الكلام، وذلك بتنظيفها باستخدام كود فلتر البيانات "Keyword Filtering" من: {علامات الترقيم، الأرقام، الحروف والأدوات، الأفعال}، للحصول على البيانات الاسمية، على افتراض أن المصطلحات التقنية هي مصطلحات أو مركبات اسمية، مع تصحيح الأخطاء التي وقعت أثناء التحليل السابق، ثم نحدد العلاقات النحوية بين المفردات حسب أقسام الكلام التي تنتمي إليها.

وتشمل المخرجات: البيانات الاسمية فقط (الأسماء والصفات)، ولأن المدونة مكتوبة بلغة تقنية خالصة باعتبارها دليل استخدام لجهاز الطابعة، فإن الكلمات والمفردات الناتجة، غالبا ما ستخلو من ألفاظ اللغة العامة، أي أنها ستترواح بين وجود: الكلمات المفاتيح والمصطلحات، وبعض مفردات اللغة العامة المرتبطة بلغة الكتابة التقنية.

**بعد تنظيف البيانات باستخدام كود فلتر البيانات "Keyword Filtering" مع حذف بيانات التردد والجذور لاحظنا أنه لا يزال من الصعب الاعتماد الكلي على الآلة في استخراج مصطلحات مجال ما، إذ أنّها تقوم كما نلاحظ في الجدول الاتي استخراج معظم الكلمات الاسمية باعتبارها**

مصطلحات أو كلمات مفتاحية، لذا لا يمكن للمصطلحي الاستغناء عن التدخل اليدوي وتصفية القوائم المصطلحية قدر الإمكان؛ لأن معظم الخوارزميات أو برامج استخراج المصطلحات تعتمد على الإحصاء وحساب تردد الكلمات في النص دون التمييز الواضح بينها، كما يتم الاعتماد على قوائم مصطلحات يتم إضافتها إلى شيفرة العمل أثناء البرمجة.

## الجدول 2: يوضح نتائج التدخل اليدوي لاستخراج المصطلحات غير المركبة.

| المخرجات | النوع النحوي | المصطلح | المخرجات |
|----------|--------------|---------|----------|
| التثبيت  | اسم          | هيرتز   | اسم      |
| الجهاز   | اسم          | توصيل   | اسم      |
| حريق     | صفة          | مقبس    | اسم      |
| إمداد    | اسم          | مأخذ    | صفة      |
| الطاقة   | اسم          | وصلة    | اسم      |
| جهد      | اسم          | سلك     | اسم      |
| فولت     | اسم          | التيار  | اسم      |

تم سابقا استخراج المصطلحات ذات الكيان الواحد، سنقوم باستخراج المصطلحات المركبة بالاعتماد على العلاقات بين الفئات النحوية، كما أشرنا إلى ذلك سابقا، حيث أن التركيبات المصطلحي في العربية معقدة بعض الشيء، لاعتبارات أهمها النقل والترجمة، إضافة إلى خصوصية بعض المجالات التي تفرض فيها تركيبات لغوية اصطلاحية معقدة على سبيل المثال لا الحصر: ارتفاع درجة حرارة الغليان، السرعة الجزيئية الأكثر احتمالا... وغيرها<sup>1</sup>، سنكتفي في هذا العنصر باستخراج التراكيب البسيطة غير المعقدة؛ مثل<sup>2</sup>: -اسم ونعت. -اسم واسم آخر. -اسم وحرف جر واسم آخر. -اسم وحرف جر ومحدد واسم آخر. -اسم وحرف جر وفعل في صيغة المصدر. سنقوم بعد ذلك بتصفية القوائم وتنظيفها للحصول على مداخل مصطلحية يمكن اعتمادها في معجم تقني متخصص. ثم نقوم باستيراد أداة n-gram من مكتبة (nltk.util)، تعتمد على سلاسل الكلمات، بحيث يمكن اختيار طول السلسلة بتحديد عدد الكلمات المطلوب استخراجها من بيانات التحليل السابقة التي حصلنا عليها، ثم حذف السلاسل

<sup>1</sup>. ينظر: سماعنه جواد حسني، "التركيب المصطلحي طبيعته النظرية وأنماطه التطبيقية"، مجلة اللسان العربي، مكتب تنسيق التعريب ع: 50، 2000م، ص: 45-46.

<sup>2</sup>. ينظر: لوم ماري -كلود، علم المصطلح مبادئ وتقنيات، تر: ربما بركة، ط1، المنظمة العربية للترجمة، بيروت، 2012م، ص: 271.

المكررة للحصول على المطلوب، نشغل كود. وبعد تشغيل الكود؛ وحذف السلاسل المكررة حصلنا على النتائج التالية:

| تثبيت الجهاز     | إمداد الطاقة | جهد كهربائي | مقبس مأخذ  |
|------------------|--------------|-------------|------------|
| صدمة كهربائية    | تردد كهربائي | وصلة تأريض  | سلك التيار |
| التيار الكهربائي |              |             |            |

الجدول 3: يوضح مخرجات البيانات اللغوية بعد تشغيل الكود وحذف السلاسل المكررة.

|               |              |                  |            |
|---------------|--------------|------------------|------------|
| تثبيت الجهاز  | حدوث حريق    | التيار الكهربائي |            |
| صدمة كهربائية | إمداد الطاقة | وصلة تأريض       | سلك التيار |
| جهد كهربائي   | تردد كهربائي | توصيل الجهاز     | مقبس مأخذ  |

الجدول 4: يوضح استخلاص المصطلحات المركبة بعد التدخل اليدوي.

### المطلب الثالث: مرحلة تمثيل المصطلحات والعلاقات دلاليا:

سأقدم في هذا العنصر نموذجا تطبيقيا بسيطا بسبب تشعب الموضوع وتعقيده؛ بحيث يشمل المثال المصطلحي الجوانب الأساسية، إضافة إلى توضيح خطوات العمل، ثم سأذكر النتائج العامة للتمثيل الدلالي للمصطلحات والعلاقات. ونحتاج فيما يأتي إلى تمثيل المصطلح "صدمة كهربائي" باستخدام إطار عمل الـ RDF وهو إحدى لغات الوصف والتمثيل في الويب الدلالي، وذلك لوصف المصطلحات التقنية وخصائصها اللغوية، ثم سنقوم بربطها بقواعد البيانات المصطلحية أو المعجمية لتوفير شبكة من البيانات الدلالية قابلة لاسترجاع المعلومات.

### أولا: تحديد المصطلح وعلاقاته التي نرغب بتمثيلها مبدئيا:

المصطلح: صدمة كهربائية

- نوع من: الصعق الكهربائي، حرق كهربائي.
- بسبب: التيار الكهربائي، موقع غير آمن، شدة الجهد الكهربائي، قوة التيار الكهربائي.
- تؤدي إلى: الإصابة بحروق، الموت.
- الوقاية: عزل الجسم، ترك مسافة.

أثر استخدام تقنيات معالجة اللغة الطبيعية للويب الدلالي في صناعة المعاجم المتخصصة الآلية: دراسة تطبيقية في المصطلحات التقنية.

باستخدام لغة بايثون ومكتبات (rdflib, SPARQLWRAPPER, spaCy, nltk)، نقوم بتمثيل هذه البيانات بالصيغة التالية:

```
# إنشاء Graph
g = Graph()

# تعريف Namespaces
EX = Namespace("http://example.org/ontology#")
RDFS = Namespace("http://www.w3.org/2000/01/rdf-schema#")
DC = Namespace("http://purl.org/dc/elements/1.1/")
```

ثانيا: توفير معلومات إضافية تقدم وصفا للمفردات والمفاهيم: نقوم بإضافة ثلاثيات RDF المتعلقة بتسمية المصلح في اللغة العربية: "صدمة كهربائية. @ar"، وترجمته الإنجليزية "Electric Shock. @en"، مع إضافة المفهوم كما هو واضح في الصيغة التالية :

```
# إضافة تلميحات RDF
g.add((EX.صدمة_كهربائية, RDF.type, EX.(مصطلح)))
g.add((EX.صدمة_كهربائية, RDFS.label, Literal("صدمة كهربائية",
lang="ar"))))
g.add((EX.صدمة_كهربائية, RDFS.label, Literal("Electric Shock",
lang="en"))))
g.add((EX.صدمة_كهربائية, RDFS.comment, Literal(
الصدمة الكهربائية تحدث الصدمة الكهربائية عند حدوث اتصال بين أي
جزء من الجسم مع مصدر كهربائي مما يسبب مرور تيار كهربائي كافٍ من خلال
الجلد، العضلات، أو الشعر (في الإنسان)، عادةً نستخدم تعبير «الصدمة
الكهربائية» لوصف التضرر الذي للكهرباء
lang="ar"))))
```

وفي هذه الصيغة قمنا بإضافة الفئة النحوية التي ينتمي إليها المصطلح:

```
# إضافة الفئة النحوية للمصطلح "صدمة كهربائية"
g.add((EX.كهربائية, EX.فئة_نحوية, Literal("Noun+Adj")))

# إضافة تعليق حول الفئة النحوية
g.add((EX.صدمة_كهربائية, RDFS.comment, Literal("الفئة النحوية  
للمصطلح هي اسم وصفة", lang="ar")))
```

تساعدنا الخاصية التالية على ربط المصطلح بالموارد الرقمية المتاحة في الانترنت التي تتضمن وصفاً، أو معلومات سواء كانت مقالات أم صوراً...إلخ عن المدخل المصطلحي لدينا، وتوفير شبكة دلالية مترابطة له عبر الانترنت.

```
# إضافة المصادر
g.add((EX, صدمة_كهربائية, DC.source,
URIRef("https://ar.wikipedia.org/wiki/صدمة_كهربائية")))
g.add((EX, صدمة_كهربائية, DC.source,
URIRef("https://www.msmanuals.com/ar/home_
الإصابات_والْقِسْم_الإصابات_الناجمة_عن-
الصناعات_الكهربائية/السواعق
")))
g.add((EX, صدمة_كهربائية, DC.source,
URIRef("https://siwar.ksaa.gov.sa/result-page-public/
search?searchType=lemma&start=20%صدمة_كهربائية")))

```

**ثالثا: إثراء المصطلحات بالمعلومات الوصفية:** يشير الرمز (URIREF) إلى المرجع الرقمي أو المصدر الذي تتوفر فيه معلومات عن المصطلح. ثم في النهاية قمنا بإثراء المصطلح بعلاقات متنوعة تربطه بمصطلحات أو مفاهيم أخرى مثل: (مصطلحات لها نفس السياق، الأضرار، المجال والفئة اللذان تنتمي إليهما).

```
# إضافة العلاقات
g.add((EX, نوع_من, EX, صدمة_كهربائية. EX))
g.add((EX, نوع_من, EX, صدمة_كهربائية. EX))
g.add((EX, تيار_كهربائي. EX, بسبب, EX, صدمة_كهربائية. EX))
g.add((EX, إصابة_بحروق. EX, تؤدي_إلى, EX, صدمة_كهربائية. EX))
g.add((EX, مصطلح_طبي. EX, مجال, EX, صدمة_كهربائية. EX))
g.add((EX, الإصابات_والتسمم. EX, فئة, EX, صدمة_كهربائية. EX))
```

**رابعا: التأكد من قدرتنا على استرجاع المعلومات:** يمكن تنفيذ استعلامات SPARQL للتأكد من قدرتنا على استرجاع المعلومات من ملفات RDF التي تم تخزينها، وإجراء بحث حول أي جزء من المعلومات المدخلة. وبعد تنفيذ الاستعلام حصلنا على النتائج التالية:

```
Label: الصدمة الكهربائية تحدث الصدمة, Comment: الصدمة الكهربائية تحدث الصدمة
الكهربائية عند حدوث اتصال بين أي جزء من الجسم مع مصدر كهربائي مما يسبب
مرور تيار كهربائي كافٍ من خلال الجلد، العضلات، أو الشعر (في الإنسان). عادةً
نستخدم تعبير «الصدمة الكهربائية» لوصف التعرض للكهرباء.
Label: Electric Shock, Comment: الصدمة الكهربائية تحدث الصدمة
الكهربائية عند حدوث اتصال بين أي جزء من الجسم مع مصدر كهربائي مما
يسبب مرور تيار كهربائي كافٍ من خلال الجلد، العضلات، أو الشعر (في الإنسان).
عادةً نستخدم تعبير «الصدمة الكهربائية» لوصف التعرض للكهرباء.
Label: الفئة النحوية للمصطلح هي اسم و, Comment: الصدمة الكهربائية
صفة.
Label: Electric Shock, Comment: الفئة النحوية للمصطلح
هي اسم و صفة.
```

## الخاتمة:

ركزنا في هذه الدراسة على كيفية تنفيذ واستخدام تقنيات معالجة اللغة الطبيعية للويب الدلالي في مجال المصطلحات التقنية، وتوظيف مخرجاتها في بناء معجم آلي متخصص، من خلال استخلاص المعلومات المصطلحية من المدونات النصية الرقمية، وتحليلها صرفيا ونحويا، وربطها بسياقاتها. كما يمكن الويب الدلالي من تمثيل العلاقات الدلالية بين المصطلحات ومفاهيمها باستخدام البيانات المهيكلة والأنطولوجيات، ما يساهم في فهم خصائص المصطلحات داخل النصوص.

ويمكن استغلال هذه النتائج في تطوير معاجم متخصصة قابلة للتوسيع والتحديث. مع إمكانية استرجاع المعلومات المرتبطة بالمصطلحات من خلال تنفيذ استعلامات SPARQL لاستخراج المعلومات من مخازن الـ RDF، كما تتيح أنظمة العلاقات (سواء المفاهيمية، أو الهرمية) المستخرجة في تصنيف المصطلحات وتمثيلها بيانيا، مع الأخذ في الاعتبار احتياجات المستخدمين.

كما يعتبر الاستثمار في مجالات معالجة اللغة الطبيعية والويب الدلالي مهما لتطوير قواعد البيانات المصطلحية في مجالات مختلفة، وتوفير واجهات برمجة تطبيقات (API) التي يمكن استعمالها كمكتبة برمجة داعمة أثناء التحليل الآلي للنصوص واستخراج المصطلحات، لتسهيل عمل المطورين واللغويين في تطوير أنظمة المصطلحات وبناء المعاجم الآلية المتخصصة، والتغلب على ندرة الموارد المصطلحية الرقمية، وتعزيز العمل بالمعاجم واستخدام اللغة العربية.

### قائمة المصادر والمراجع:

#### أولاً: الكتب

1. جاد الكريم عبد الله، عبقرية العربية والذكاء الاصطناعي - آفاق وطموحات -، ط1، دار النابغة للنشر والتوزيع، طنطا، 2020م.
2. الفيفي عبد الله بن يحيى، مدخل إلى اللسانيات الحاسوبية، ط1، مركز الملك عبد الله الدولي لخدمة اللغة العربية، الرياض، 2017م.
3. اللتيني يحيى أحمد، أسس صناعة المعجم المحوسب، ط1، كنوز المعرفة، الأردن، 2019م.
4. لوم ماري -كلود، علم المصطلح مبادئ وتقنيات، تر: ريم بركة، ط1، المنظمة العربية للترجمة، بيروت، 2012م.
5. ماينارد ديانا وآخرون، معالجة اللغات الطبيعية للويب الدلالي، تر: خالد بن عبد الرحمن الميمان، ط1، مركز الملك عبد الله بن عبد العزيز الدولي لخدمة اللغة العربية، الرياض، 2019م.
6. يسري مصطفى وآخرون، الطاقة الشمسية: النظرية والتطبيق، ط1، مؤسسة الأمة العربية، القاهرة، 2020.
7. Baader, Franz, and al. THE DESCRIPTION LOGIC HANDBOOK: Theory, implementation, and applications. Cambridge: Cambridge university press. 2017.
8. Hobson Lane and al., Natural Language Processing in Action, 2nd edition, Manning Publications, New York, 2021.

#### ثانياً: المقالات

1. أكماضان محمد، "لمحة عامة عن الأسس النظرية والمنهجية للعمل المصطلحي"، دائرة الترجمة العربية، نيويورك، 2015م.

2. سمانه جواد حسني، "التركيب المصطلحي طبيعته النظرية وأنماطه التطبيقية"، مجلة اللسان العربي، مكتب تنسيق التعريب ع: 50، 2000م.
3. العنزي محمد بن نافع، "المصاحبة اللفظية في العربية المعاصرة وأثرها في تغير الدلالة"، مجلة الدراسات اللغوية، مركز الملك فيصل للبحوث والدراسات الإسلامية، م15 - ع01، 2012م.
4. Bellandi Andrea. "Le Risorse Linguistiche nell'era del Web Semantico". AIDAinformazioni, Cacucci Editore, Vol. 40, no. 1-2.
5. Berners-Lee, Tim, and al. "The Semantic Web". Scientific American, Scientific American Inc., Vol. 284, No. 5, 2001.
6. Jarrar Mustafa, and al. "ALMA: Fast Lemmatizer and POS Tagger for Arabic". Procedia Computer Science, 244, 2024.

### ثالثاً: المواقع الإلكترونية:

1. Harris, steve, and Seaborne Andy, "SPARQL 1.1 Query Language", W3C Recommendation, 2013, Available at: <https://www.w3.org/TR/sparql11-query/>
2. <https://oip.manual.canon/USRMA-3694-zz-SSar/contents/01020000.html>
3. <https://sina.birzeit.edu/qabas/lemma/202007869>
4. <https://sina.birzeit.edu/sinatools/documentation/index.html>.
5. Schreiber, Guus, and Yves Raimond. "RDF 1.1 Primer". W3C, 2014, Available at: <https://www.w3.org/TR/rdf11-primer/>.
6. Tim Berners-Lee, "linked data", 2009, Available at: <https://www.w3.org/DesignIssues/LinkedData>, Accessed 11 August 2023.
7. W3C. "OWL 2 Web Ontology Language: Overview". 2012, Available at: <https://www.w3.org/TR/owl2-overview/#ref-owl-2-specification>. Accessed 11 August 2023.