11. Elkateb, S., Black W., Vossen P., Farwell D., Rodríguez H., Pease A., Alkhalifa M. “Arabic WordNet and the Challenges of Arabic”. In proceedings of Arabic NLP/MT Conference, London, U.K, 2006.

12. Mustafa Jarrar, Building a Formal Arabic Ontology, In proceedings of the Experts Meeting on Arabic Ontologies and Semantic Networks. Alecso, Arab League. Tunis, July 26-28, 2011.

13. Attia, M., Rashwan, M., Ragheb, A., Al-Badrashiny, M., Al-Basoumy, H., Abdou, S., A Compact Arabic Lexical Semantics Language Resource Based on the Theory of Semantic Fields, Lecture Notes on Computer Science (LNCS): Advances in Natural Language Processing, Springer-Verlag Berlin Heidelberg www.SpringerOnline.com, LNCS/LNAI; Vol. No.

المحتوى الرقمي العربي. كما لَفتْنا الانتباه إلى ضرورة تطوير وإنجاز الانطولوجيات العربية لما لها من أهمية قصوى في تطوير البرمجيات اللغوية المُحوْسَبة.

## المراجع


1. موقع الباحث العربي www.baheth.info
2. قاموس ومعجم المعاني www.almaany.com
3. المعاجم www.maajim.com
4. معجم الوسيط الإلكتروني

http://sibawayh.emi.ac.ma/safar/applications

5. معجم اللغة العربية التفاعلي

https://sourceforge.net/projects/almuajam

6. عبد المجيد بن حمادو، المعجم العربي الإلكتروني: أهميته وطرق بنائه، 2011

7. حسام الخطيب، مستقبل المعاجم العربية في الجمع بين الرقمي والورقي، http://www.startimes.com/?t=22926221

8. أنور الجمعاوي، المعجم العربي الإلكتروني المختص : قراءة نقدية في نماذج مختارة http://www.aot.org.lb/Attachments/Attachment88_120.pdf

9. عز الدين غازي، بناء المعاجم الالكترونية أم تقييس الكفاية المعجمية على الحاسوب، الحوار المتمدن–العدد: 2292 - 2008 / 5 / 25.

10. عبد القادر الغنامي، إحصاء المعاجم العربية

http://www.atida.org/forums/showthread.php?t=942

الدلالة فيتمثل من خلال تعريف المصطلح،مع الأخذ بعين الاعتبار بأن المصطلح متصل بمصطلحات أخرى عبر علاقات دلالية، وبالتالي يحمل التعريف نصا دلاليا تتم صياغته بطريقة منطقية لتمكين الحاسوب من استخدامه. فالتعريف ليس فقط موجه للقراءة ولكنه مصاغ بطريقة منطقية يمكن استخدامه من طرف الحاسوب. هكذا عندما نعرف بمصطلح "الكمر" بالنص الآتي "من البسر: ما لم يرطب على نخله، ولكنه سقط فأرطب على الأرض"، نستنتج حاسوبيا (مع تدقيق لغوي بعدي ) أن الكمر له علاقة دلالية مع مصطلح البسر وهي علاقة النوعية، فيمكن بالتالي برمجة الحاسوب على أن يستنتج علاقة النوعية ما بين كل مصطلح "x" و "y" كلما عُرِفَ"x"في المعجم ب "من y ...."

يشهد ميدان أنطلوجيات اللغة العربية شحا وخصاصا كبيرا إلا أن هذا لا ينفي وجود مجهودات من قبل الباحثين لتطوير هذا الميدان. ونذكر بأهم محاولات بناء انطولوجيات اللغة العربية: شبكة الكلمات العربية (وردنيت العربي) **[12]**ArabicWordNet وأنطولوجيا اللغة العربية لجامعة بيرزيت بفلسطين**[13]** وأنطولوجيا شركة RDI**[14]**، بالرغم من وجود مشاريع بعضها قيد الانجاز والبعض الآخر مجرد نوايا. إجمالا، ليس هناك مشروع متكامل وعملي يمكن استغلاله في إنشاء تطبيقات لغوية حاسوبية، اللهم مشروع Arabic WordNet وهو عمل غير مكتمل لأنه يغطي فقط جزءا بسيطا من نظام اللغة العربية، والأهم من هذا فهو غير نابع من خصوصيات اللغة العربية.

**5. خلاصة**

تطرقنا في هذه الورقة إلى مفهوم المعاجم الرقمية بصفة عامة وعرفنا بآليات وأدوات تحويل المعجم الورقي إلى معجم رقمي، كما أشرنا إلى أهمية تطويرها وتقييسها ونشرها عبر مختلف وسائل التكنولوجيا المعاصرة بهدف إثراء

الشكل رقم 3 هو مثال لقسط صغير لما يجب أن يكون عليه المعجم الرقمي في صيغته النهائية. بالموازاة مع تصميم وتقييس الموارد المعجمية يتم العمل على تصميم وتطوير واجهة استغلال المعجم والتي يجب أن تتميز بمجموعة من الخصائص كسُهُولة الاستعمال، وإمكانية تثبيت البرنامج على أجهزة مختلفة من حواسب وهواتف ذكية وشبكة الأنترنيت، وتقديم إمكانية بحث متنوع يلبيّ حاجيات المستعمل.

**4. التحديات والآفاق المستقبلية**

لقد شهد التاريخ تطوير عدد هائل من المعاجم للغة العربية منها ما هو مُرقمن، ومنها ما ينصح برقمنته حسب الخطة المقترحة في الفقرة 3 لتعميم الفائدة، وتداول المعجم عبر وسائل الاتصال الجديدة. لكن هذه المعاجم، تبقى رغم حمولتها وكثرتها، غير قادرة على مُواكبة حاجيات البرمجيات الحديثة من محركات البحث والترجمة الآلية، إلخ. والتي تتمثل في موارد مناسبة لا تقتصر فقط، على الجانب المصطلحي وإنما تمتد لتشمل الجانب الدلالي.

أغلب تطبيقات هندسة اللغة من محركات البحث وأنظمة الترجمة في حاجة ماسة إلى نوع جديد من المعاجم التي تحظى بتغطية (بتمثيلية) على المستوى الدلالي؛ ويصطلح على هذه المعاجم بشبكات الكلمات الدلالية أو الأنطولوجيات. السبب الرئيسي هو أن هذه البرامج أصبحت تحتاج إلى استعمال المستوى الدلالي في خوارزمياتها. تخزن وتُمثل الدلالة في شبكة الكلمات (الأنطولوجيات) على مستويين: المستوى الأول يهمّ ربط المصطلحات أو بالأحرى المفاهيم فيما بينها بعلاقات دلالية؛ فمثلا تضاف علاقة الترادف بين مصطلحي القطّة والهرّة، وعلاقة النوعية بين القطة والحيوان، أو علاقة الجزئية بين الرأس والقطة. هكذا، تمثل الأنطولوجيا جميع مصطلحات المعجم بالإضافة إلى العلاقات الدلالية فيما بينها لتُكوِن شبكة كلمات. أما المستوى الثاني من

**الشكل 3 :مثال جزء من معجم موصوف بمعيار LMF**

```
<?xml version='1.0' encoding='UTF-8'?>
<Lexicon>
        <feat att='language' val='arab' />
        <LexicalEntry id='مائم' >
                <feat att='partOfSpeech' val='noun' />
                <Lemma>
                        <feat att='writtenForm' val='مائم' />
                        <feat att='scheme' val='لَنْجَم' />
                </Lemma>
                <WordForm>
                        <feat att='wittenForm' val='بَنْأم' />
                        <feat att='grammaticalNumber' val='singular'/>
                </WordForm>
                <WordForm>
                        <feat att='WittenForm'  val='  مائمة' />
                        <feat att='grammaticalGender' val='feminine'/>
                        <feat att='grammaticalNumber' val='singular'/>
                </WordForm>
                <WordForm>
                        <feat att='WittenForm'  val='  بمآت' />
                        <feat att='grammaticalNumber' val='plural'/>
                </WordForm>
                <WordForm>
                        <feat att='WittenForm'  val='   بمآت' />
                        <feat att='grammaticalGender' val='feminine'/>
                        <feat att='grammaticalNumber' val='plural'/>
                </WordForm>
                <RelatedForm targets='م أ ت ' >
                        <feat att='type' val='root' />
                </RelatedForm>
```

بعد إنهاء مهمة إنشاء قاعدة البيانات المعجمية يستلزم تقييسها أي كتابتها بطريقة معيارية متعارف عليها دوليا. هذه العملية تمكن من تيسير عملية تبادل المعاجم بين الأشخاص والمؤسسات قصد الإثراء المشترك باستغلال البنية الموحدة كما تُمكن من دمج معطياتها في التطبيقات اللغوية الحاسوبية. أكثر المعايير تداولا بالنسبة للموارد المعجمية هو معيار LMF "إطار التنميط اللغوي[6] Lexical Markup Framework"وهو مقياس إيزو"ISO" صدر سنة 2008 ولا يزال يخضع للتطوير، ويقوم على تحديد البنية العامة للمدخلات اللغوية وكيفية تخزينها رقميا حتى يسهل استعمالها في عدة تطبيقات مختلفة.

6 - https://ar.wikipedia.org/wiki/هيكلـ الترميزـاللغوي

لكن التجربة أفضت إلى تفضيل منهجية نصف آلية نعرضها في هذا المقال مقارنة مع المنهجية الآلية. بعد مهمة التدقيق يتم تحويل النصوص المحصل عليها إلى قاعدة بيانات تمثل نسخة مماثلة للمعجم الورقي.

يقرأ البرنامج الحاسوبي النص المحصل عليه بعد تدقيقه، ويبحث عن معلومات مفصلة فيضعها في قاعدة بيانات المعجم الرقمي، وهذه بيانات مبنية من خلال:

- المصطلح العربي: المصطلح الأصلي باللغة العربية.

- المصطلح المقابل باللغات الأخرى: ترجمة المصطلح الأصلي إلى اللغات الأخرى وخاصة الإنجليزية التي تحظى بتداول كبير في المجالات العلمية.

- الجذر: جذر المصطلح العربي وينصح كتابته بإضافة فراغ بين الحروف (حتى لا يختلط مع صيغة الفعل في الماضي).

- قسم الكلمة: في غالبية الأحيان يمكن الاقتصار على الأقسام الثلاثة: الحرف والفعل والاسم.

- الوزن: وزن المصطلح.

- الجنس: ذكر أو مؤنث بالنسبة للاسم.

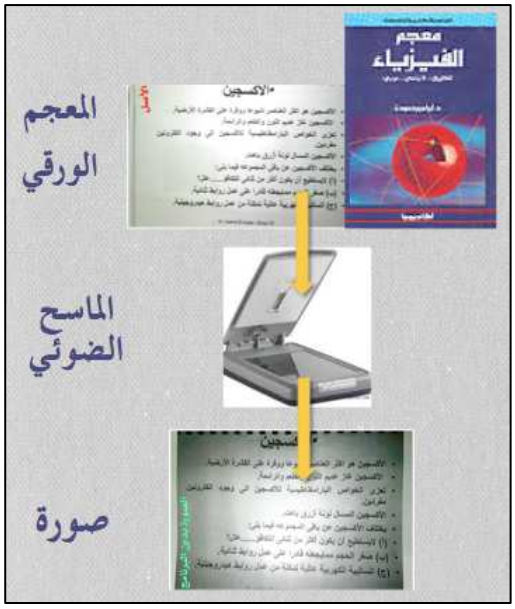
- الجمع: خصوصا جمع التكسير.

- ميدان الاستعمال: إضافة الميادين التي يمكن أن يستعمل فيها المصطلح؛ مثلا مصطلح "إبرة" يستعمل في ميدان الخياطة والطّب.

- التعريف الدلالي: نص قصير يشرح ويعرف بالمصطلح، وسنعود إلى هذا المفهوم لاحقا لما له من أهمية وعلاقة بصناعة الأنطولوجيات أو المعاجم الدلالية.

تعرف دقة وجودة برامج التعرف الضوئي على الحروف العربية مستوى أقل من نظيراتها في اللغات اللاتينية التي تفوق 90% من الدقة،إضافة إلى ذلك فعدد هذه البرمجيات لا زال محدودا، بعضها مجانية ومفتوحة المصدر وبعضها الآخر تجارية. نذكر على سبيل المثال برنامج القارئ الآلي لشركة صخر[5]. ويرجع سبب انخفاض نسبة الدقة بالأساس إلى خصوصيات الخط العربي الذي يسمح بكتابة كل حرف بطرق متعددة تقبل التطويل والتفنن والزيادة إلخ، تجعل البرنامج الحاسوبي لا يعرف مسبقا كل الحالات التي سيتواجد فيها حرف معين ويصعب بالتالي التعرف عليه.

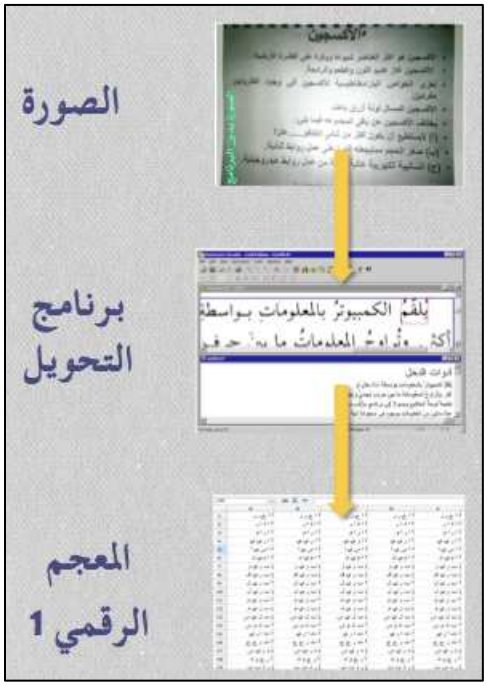
**الشكل 2 : خطوات المرحلة الثانية**




تواضع دقة برنامج التّعرف الضوئي على الحروف العربية تحتّم إضافة مهمة ضرورية في نهاية المرحلة الأولى عبر إنجاز التدقيق اليدوي من أجل صحة النصوص المحصل عليها. هذه الإضافة في الجهد والوقت تجعل الباحث يفكر في تغيير المنهجية بمسك المعجم يدويا بالكامل عوض اللّجوء إلى برامج حاسوبية.

5 - http://www.sakhr.com/index.php/en/solutions/ocr

**الشكل 1 :خطوات المرحلة الأولى**




كما يوضح الشكل رقم 1، فإن المرحلة الأولى تنطلق من المعجم الورقي الذي يتم مَسْحه كاملا باستعمال ماسح ضوئي مرفوق ببرنامج حاسوبي يساعد على الحصول على الصورة المقابلة لكل صفحة من المعجم، بالتفاصيل المرغوب فيها من دقة الصورة والألوان والحجم الخ. أما فيما يخص المرحلة الثانية وكما يوضح الشكل رقم 2، فتنطلق من الصّور المحصل عليها سابقا حيث يتم اللجوء إلى برنامج "التعرف الضوئي على الحروف العربية" والذي تتركز وظيفته على تقطيع كل صورة إلى سطور، ثم تقطيع صورة السّطر الواحد إلى صور أصغر تمثل كل واحدة فيها كلمة، وتقسم الصورة الصغيرة للكلمة إلى أجزاء يقوم النظام بالتعرف على الحرف العربي المناسب لكل جزء. وبطريقة معكوسة يقوم النظام بتجميع الكلمات انطلاقا من الحروف المحصل عليها، ثم السطر فالنص المماثل للصفحة الكاملة.