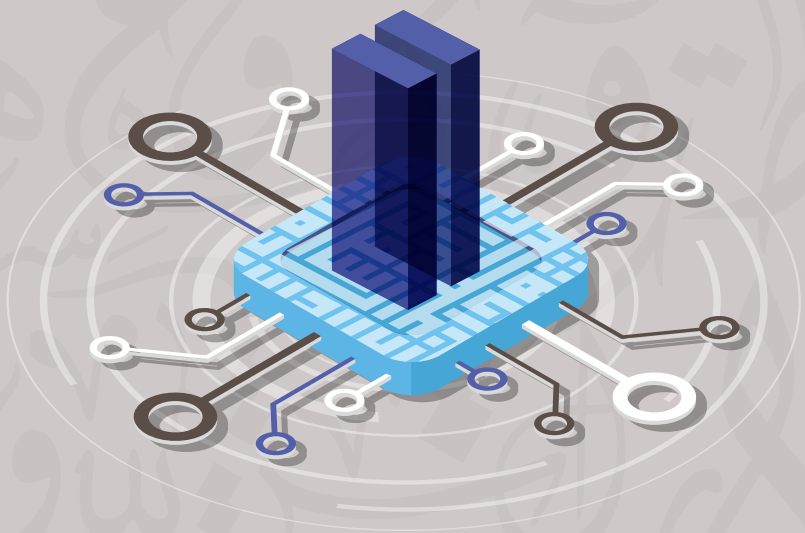
مجمع الملك سلمان العالمي للغة العربية
King Salman Global Academy for Arabic Language

# معايير الحوسبة اللغوية العربية



١٠

السياسات والمعايير

مجمع الملك سلمان العالمي للغة العربية
King Salman Global Academy for Arabic Language

# معايير الحوسبة اللغوية العربية

١٠

السياسات والمعايير




**معايير الحوسبة اللغوية العربية**

الطبعة الأولى
١٤٤٧هـ ٢٠٢٥م

البريد الإلكتروني: nashr@ksaa.gov.sa

ح/ مجمع الملك سلمان العالمي للغة العربية، ١٤٤٧هـ
معايير الحوسبة اللغوية العربية. / مجمع الملك سلمان العالمي للغة العربية. - الرياض، ١٤٤٧هـ

٣٤٠ ص؛ ١٧ × ٢٤ سم - (السياسات والمعايير؛ ١٠)

رقم الإيداع: ١٤٤٧/٥٩٤٧
ردمك: ٧-٤٨-٨٥٢٥-٦٠٣-٩٧٨

لا يسمح بإعادة إصدار هذا الكتاب، أو نقله في أي شكل أو وسيلة، سواء أكانت إلكترونية أم يدوية، بما في ذلك جميع أنواع تصوير المستندات بالنسخ، أو التسجيل أو التخزين، أو أنظمة الاسترجاع، دون إذن خطي من المجمع بذلك.

الآراء الواردة في هذا الكتاب تمثِّلُ رأيَ المؤلف، ولا تعكسُ - بالضرورة - رأي المجمع.

**هذه الطبعة إهداء من المجمع ولا يُسمح بنشرها ورقياً أو تداولها تجارياً**


بسم الله الرحمن الرحيم

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

# محتويات الكتاب


| الموضوع | الصفحة |
|---|---|
| مقدمة المجمع | ٩ |
| قائمة المصطلحات | ١١ |
| **الفصل الأول:** وثيقة معايير الحوسبة اللغوية العربية | ١٥ |
| مقدمة | ١٧ |
| وثائق معايير الحوسبة اللغوية العربية | ٢٠ |
| الأهمية والمسوغات | ٢١ |
| تحديات معايير حوسبة اللغة العربية | ٢٤ |
| منهجية بناء معايير الحوسبة اللغوية العربية وتطويرها | ٢٨ |
| جهات تطوير معايير الحوسبة اللغوية | ٣١ |
| معايير بناء المعاجم الإلكترونية العامة والمختصة | ٣٥ |
| معايير بناء منصات تعليم اللغة العربية لغةً ثانيةً | ٤٠ |
| معايير بناء وتوسيم المدونات اللغوية | ٤٤ |
| معايير دعم اللغة العربية في التطبيقات الحاسوبية | ٤٧ |
| المراجع | ٥٠ |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً


| الموضوع | الصفحة |
|---|---|
| **الفصل الثاني:** معايير بناء المنصات الرقمية لتعليم اللغة العربية لغة ثانية | ٥٥ |
| **ملخص** | ٥٧ |
| **مقدمة** | ٥٧ |
| **مدخل تعريفي** | ٦٠ |
| **الدراسات السابقة** | ٦٧ |
| **نطاق المعايير وحدودها** | ٧١ |
| **قائمة المعايير** | ٧٢ |
| **المراجع** | ١٠٦ |
| **الفصل الثالث:** معايير بناء المعاجم الرقمية العامة | ١١١ |
| **ملخص** | ١١٣ |
| **مقدمة** | ١١٣ |
| **مدخل تعريفي** | ١١٦ |
| **الدراسات السابقة** | ١٢٠ |
| **نطاق المعايير وحدودها** | ١٣٣ |
| **قائمة المعايير** | ١٣٤ |
| **المراجع** | ١٥٨ |
| **الفصل الرابع:** معايير بناء المعاجم المختصة | ١٦١ |
| **ملخص** | ١٦٣ |
| **مقدمة** | ١٦٣ |
| **مدخل تعريفي** | ١٦٥ |
| **الدراسات السابقة** | ١٧٠ |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً


| الموضوع | الصفحة |
|---|---|
| نطاق المعاييروحدودها | ١٨٠ |
| قائمة المعايير | ١٨١ |
| المراجع | ٢٠٧ |
| **الفصل الخامس:** معايير بناء المدونات اللُّغوية | ٢١١ |
| ملخص | ٢١٣ |
| مقدمة | ٢١٣ |
| مدخل تعريفي إلى بناء المدونات اللُّغوية | ٢١٥ |
| مسوغات وضع المعاييرفي بناء المدونات اللُّغوية | ٢٢٣ |
| الدراسات السابقة | ٢٢٨ |
| نطاق المعاييروحدودها | ٢٣٩ |
| قائمة المعايير | ٢٣٩ |
| المراجع | ٢٤٩ |
| **الفصل السادس:** توسيم المدونات اللُّغوية العربية | ٢٥١ |
| ملخص | ٢٥٣ |
| مقدمة | ٢٥٣ |
| مدخل تعريفي إلى توسيم المدونات اللُّغوية العربية | ٢٥٥ |
| مُسوِّغات بناء المعاييرلتوسيم المدونات اللُّغوية العربية | ٢٦٤ |
| الدراسات السابقة | ٢٦٧ |
| نطاق المعاييروحدودها | ٢٧٤ |
| قائمة المعايير | ٢٧٥ |
| المراجع | ٣٠٢ |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً


| الموضوع | الصفحة |
|---|---|
| **الفصل السابع:** دعم اللُّغة العربية في التطبيقات الحاسوبية | ٣٠٥ |
| **ملخص** | ٣٠٧ |
| **مقدمة** | ٣٠٧ |
| **مدخل تعريفي إلى دعم اللُّغة العربية في التطبيقات الحاسوبية** | ٣٠٨ |
| **مُسوِّغات بناء المعايير لدعم اللُّغة العربية في التطبيقات الحاسوبية** | ٣٠٩ |
| **الدراسات السابقة** | ٣١١ |
| **نطاق المعايير وحدودها** | ٣١٩ |
| **قائمة المعايير** | ٣٢٠ |
| **المراجع** | ٣٣٦ |
| **شكر وتقدير** | ٣٣٩ |

هذه الطبعة **إهـداء من المجمـع**
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

# مقدمة المجمع

**بسم الله الرحمن الرحيم**

تشهد العقود الأخيرة ثورةً معرفيةً وتقنيةً غير مسبوقة؛ إذ أصبحت تقنيات الذكاء الاصطناعي جزءًا أساسًا من صناعة المعرفة والاقتصاد الرقمي.

ولأن اللغة العربية وعاء فكر وثقافة، ومعرفة وحضارة تمتد لأكثر من أربعة عشر قرنًا، ونظرًا إلى تميز اللغة العربية بثراء صرفي ونحوي ودلالي، إضافةً إلى كثرة البرامج والتطبيقات؛ فإن التباين بين الباحثين والمبرمجين متوقع في اعتماد طرق متباينة من ترميز الحروف، والتشكيل، والتجذير، والتعامل مع الرسوم الكتابية؛ لذلك أصبحت حوسبتها بدقة وتوحيدها ضرورةً ملحّةً؛ لضمان حضورها الفاعل في العصر الرقمي، ومن هنا تتجلى أهمية بناء معايير موحدة لحوسبة اللغة العربية.

إن بناء معايير حوسبة اللغة يمكن أن يُسهم في قدرة الحاسوب على التعامل بفاعلية مع اللغة العربية في مجموعة من المسارات، مثل: التعليم الرقمي، وبناء المعاجم الرقمية، وبناء المدونات الرقمية، وتوسيم المدونات الرقمية، وأنظمة الترجمة الآلية، والبحث الدلالي في النصوص، والتعرف الآلي على الكلام ونحو ذلك، ويمثّل استثمارًا اقتصاديًّا وحضاريًّا مهمًّا، يمكّن من إدماج مئات الملايين من الناطقين بالعربية في الاقتصاد العالمي المعرفي الرقمي، ويضمن أن يبقى المحتوى العربي صحيحًا.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

وقد عمل المجمع على بناء (معايير حوسبة اللغة العربية)؛ لدعم حضور العربية في المجال الرقمي، وإدماج العربية والناطقين بها في مجالات الحياة الحيوية، وينتظم هذا العمل ضمن سلسلة إصدارات المجمع في مجال السياسات والمعايير اللغوية، التي نأمل أن تثري الميدان اللغوي، وأن تسهم في تعزيز المعرفة اللغوية.

ويأتي هذا العمل ضمن الدعم الدائم والتوجيه المستمر الذي تجده أعمال المجمع ومبادراته، ومشروعاته، وإستراتيجيته من سمو وزير الثقافة رئيس مجلس أمناء المجمع، الأمير/ بدر بن عبدالله بن فرحان آل سعود -حفظه الله-.

ويسعد فريق المجمع بتلقي الأفكار الإثرائية التي تجوّد العمل وتعالج أي فجوات تطرأ ضمن هذه السلسلة أو غيرها.

والله الموفق.

الأمين العام للمجمع

**أ.د. عبد الله بن صالح الوشمي**

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| العربية | الإنجليزية | التعريف |
|---|---|---|
| **اختبار القبول** | Acceptance Testing | مجموعة من الاختبارات التي تجرى لضمان تلبية التطبيق لمتطلبات المستخدمين، والعمل كما هو متوقع. |
| **النص البديل** | Alt Text | نص يوضع كبديل للصور في صفحات الويب؛ لتمكين قارئات الشاشة من قراءته. |
| **المدونة الموسَّمة** | Annotated corpus | مدونة أُضيفت إلى بياناتها وسومٌ لغوية، ولها أنواع متعددة بحسب طبيعة الوسوم. |
| **واجهة التطبيق البرمجية** | API | مجموعة من التعريفات والبروتوكولات والأدوات لبناء البرامج والتفاعل بينها برمجيا؛ لتمكن البرامج الأخرى من الوصول إلى معلومات أو خدمات معينة، بدون الحاجة لتدخل المستخدم. |
| **التطبيق** | Application | برنامج يُستخدم لإنجاز مهمة من المهام، سواء ترفيهية كانت أم احترافية، أم غيرذلك. |
| **معايير التطبيقات** | Application Standards | القواعد والإرشادات المُتَّبعة أثناء تصميم التطبيقات وتطويرها؛ لضمان الجودة والكفاءة والأمان. |
| **الذكاء الاصطناعي** | Artificial Intelligence | فرع من علوم الحاسب يُعنَى ببناء الأنظمة الذكية التي تتطلب عادة محاكاة الذكاء البشري بواسطة الآلات، وخاصة أنظمة الحاسب. |
| **البيانات الضخمة** | Big Data | مجموعات البيانات الكبيرة والمعقدة التي يصعب معالجتها باستخدام أدوات إدارة البيانات التقليدية. |
| **النمط المتعاقب** | Cascading Style Sheets – CSS | لغة توصيف تستخدم لتنسيق وتصميم مستندات الويب. |
| **التوافق** | Compatibility | قدرة التطبيق على العمل السليم الصحيح عبر مختلف الأجهزة وأنظمة التشغيل. |
| **الكشاف السياقي** | Concordance | عبارة عن قائمة بالكلمات في مدونة أو نص ضمن سياقاتها المختلفة. ويختلف الكشاف السياقي عن فهرس الكلمات في أنه يشمل معلومات أكبر، مثل تعريف الكلمة أو ترجمتها للغة أخرى، بالإضافة إلى السياقات المختلفة التي تحتوي الكلمة. |
| **المدونة اللغوية** | Corpus | عدد كبيرمن النصوص المكتوبة أو المنطوقة (الرقمية غالبًا) التي تُجمع بطرق متعددة لتمثيل استعمال لغوي عام أو خاص. |
| **توسيم المدونات** | Corpus annotation | عملية إرفاق وسوم لغوية منظّمة ببيانات المدونة (صرف، تركيب، دلالة...)، آليًا أو يدويًا. |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| العربية | الإنجليزية | التعريف |
|---|---|---|
| **أدوات التوسيم اللغوي** | Corpus annotation tools | الأدوات البرمجية المتنوعة التي تستخدم في التحليل اللغوي الآلي وأداء عملية التوسيم، وتتكون غالبًا من واجهة مستخدم رسومية لتسهيل عملية التوسيم، إضافة إلى أنظمة لإدارة عملية التوسيم اليدوي ومتابعتها، مثل إسناد مهام توسيم النصوص الرقمية إلى المشاركين، ومتابعة إنجاز مراحل التوسيم اللغوي وتنظيمه. |
| **توافقية المنصات** | Cross-Platform Compatibility | قدرة التطبيق على العمل السليم دون إخلال بالواجهة الرسومية أو المهام، عبر مختلف الأنظمة والمنصات، مثل Android، وiOS، و Windows، وغيرها. |
| **التعلم العميق** | Deep Learning | مجموعة فرعية من التعلم الآلي تستخدم الشبكات العصبية الاصطناعية Artificial Neural Networks لمعالجة البيانات وتحليلها. |
| **التشكيل** | Diacritization | إضافة الحركات وعلامات الإعـراب آليًا/يدويًا إلى النص العربي لضبط النطق والمعنى. |
| **التضمين** | Embeddings | تمثيل النصوص في شكل أرقام (متجهات) تستخدم في نماذج التعلم الآلي لمعالجة اللغات الطبيعية. |
| **الخطوط** | Fonts | أشكال وتصاميم النصوص المستخدمة في عرض الأحرف. |
| **لغة توصيف النص الفائق** | HyperText Markup Language – HTML | لغة توصيف تستخدم لإنشاء وتصميم صفحات ومواقع الويب. |
| **التكامل** | Integration | القدرة على ربط التطبيق بأنظمة أو تطبيقات أخرى لتبادل البيانات والعمل المتناسق. |
| **الموارد اللغوية** | Language Resources | مصطلح عام يشمل جميع الأوعية اللغوية التي تستخدم لحفظ وتخزين البيانات اللغوية، وتشمل المدونات اللغوية، والمعاجم والقواميس، والمسارد والشبكات الدلالية وغيرها. |
| **الإعدادات المحلية** | Locale | مجموعة من الإعـدادات المتعلقة باللغة والمنطقة الزمنية والنظام الرقمي والتاريخ والوقت والعملة. |
| **التوطين** | Localization | تكييف التطبيق ليتناسب مع اللغة والثقافة المحلية للمستخدمين. |
| **التعلم الآلي** | Machine Learning | فرع من الذكاء الاصطناعي يُعنَى بتعليم الحاسب كيفية التعلم من البيانات السابقة وتحسين الدقة بمرور الوقت وبتزايد البيانات. |
| **الترجمة الآلية** | Machine Translation | استخدام تقنيات الذكاء الاصطناعي لترجمة النصوص من لغة إلى أخرى تلقائيًا. |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| العربية | الإنجليزية | التعريف |
|---|---|---|
| **اللغة الطبيعية** | Natural Language | اللغة التي يتحدث بها البشر ويستخدمونها في التواصل اليومي، مثل العربية والإنجليزية. |
| **التنقل** | Navigation | العملية التي يستخدمها المستخدمون للانتقال بين صفحات ووظائف التطبيق. |
| **نظام التشغيل** | Operating System OS | برنامج يدير أجهزة الحاسب ويُقدّم خدمات للمستخدمين والبرامج الأخرى، مثل إدارة الملفات، وجدولة المهام، وإدارة الذاكرة. |
| **اختبار الأداء** | Performance Testing | عملية تقييم سرعة واستجابة التطبيق تحت ظروف مختلفة؛ لضمان قدرته على التعامل مع الأحمال الكبيرة. |
| **المدونة الخام** | Raw corpus | مجموعة من البيانات اللغوية بصفتها الأصيلة دون أي إضافات لغوية جوهرية عليها، منها المدونات النصية، والمدونات الصوتية أو المرئية. |
| **محاذاة من اليمين إلى اليسار** | RTL | إعداد لاتجاه عرض النصوص والواجهة من اليمين إلى اليسار، شائع في اللغات مثل العربية. |
| **قابلية التوسع** | Scalability | قدرة التطبيق على التكيف مع زيادة عدد المستخدمين أو زيادة حجم البيانات دون التأثير في الأداء. |
| **القارئ الآلي** | Screen Reader | برنامج يساعد المستخدمين المكفوفين أو ضعاف البصر على استخدام الكمبيوتر عن طريق قراءة النصوص بصوت عالٍ. |
| **الأمان** | Security | مجموعة الإجـراءات والتقنيات التي تهدف إلى حماية التطبيق والمستخدمين من الهجمات الإلكترونية والاختراقات. |
| **التقسيم** | Segmentation | عملية تقسيم النص إلى وحدات صغيرة، مثل الكلمات أو الجمل؛ لتحليلها ومعالجتها. |
| **الهاتف الذكي** | SmartPhone | الهواتف الذكية التي تتضمن أهم وظائف الحاسوب التقليدية. |
| **المدونة الصوتية** | Speech corpus | مدونة تتكون من مواد لغوية على شكل أصوات، كالمحادثات الهاتفية، والأخبار المسجلة على شكل ملفات صوتية. |
| **التفريغ الإملائي (النسخ الكتابي)** | Speech to text | تحويل المدونة الصوتية إلى مدونة نصية مكتوبة وفقا لقواعد الكتابة الإملائية. |
| **الجهاز اللوحي** | Tablet | جهاز محمول أكبر من الهاتف الذكي يعمل غالبا باللمس، ويستخدم لأغراض متعددة، مثل: القراءة والتصفح والكتابة. |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| العربية | الإنجليزية | التعريف |
|---|---|---|
| تحليل النصوص | Text Analytics | استخدام تقنيات تحليل البيانات؛ لاستخراج معلومات مفيدة من النصوص المكتوبة. |
| توسيم النصوص | Text annotation | عملية إضافة معلومات لغوية إلى مكونات النص الرقمي بطرق آلية أو يدوية، وله صور وأشكال مختلفة، ومن أشهر أنواعه: إضافة الوسوم الخاصة بأجزاء الكلام إلى النص، مثل: (الاسم، والفعل، والحرف). |
| المدونة النصية | Text corpus | مجموعة نصوص مكتوبة، منظَّمة وفق معايير اختيار وتمثيل معروفة. |
| التفريغ الصوتي (النسخ الصوتي) | Text to speech | تحويل المدونة الصوتية إلى مدونة نصية مكتوبة بالأحرف الصوتية للغة إذا كانت حروف اللغة ممثلة لأصواتها، أو باستخدام الأبجدية الدولية الصوتية IPA، دون مراعاة لقواعد الكتابة الإملائية. وقد تضاف إليها معلومات تتعلق بالمتحدثين وبيئة التحدث، مثل: جنس المتحدث ومكان التحدث (معزول صوتيا أو به ضوضاء). |
| التقطيع | Tokenisation | يقصد به تقطيع المدونة النصية إلى أجزاء أصغر، مثل تقطيع المدونة إلى مستندات، والمستندات إلى فقرات، والفقرات إلى جمل، والجمل إلى كلمات، والكلمات إلى وحدات صرفية. وذلك بحسب التطبيق المستهدف، وتكون عملية التقطيع خطوة أولى قبل التوسيم اللغوي. |
| تقطيع الكلمات | Tokenization | عملية تقسيم النص إلى أجزاء صغيرة، مثل الكلمات أو العبارات لاستخدامها في المعالجة الحاسوبية. |
| الترميز الموحد | Unicode | معيار دولي لترميز الأحرف يمكن للحاسب من خلاله تمثيل النصوص ومعالجتها في أي نظام كتابة لغوي. |
| تجربة المستخدم | User Experience – UX | الانطباع العام الذي يحصل عليه المستخدم عند استخدام التطبيق، ويتضمن سهولة الاستخدام، والتصميم، وسرعة الأداء. |
| واجهة المستخدم | User Interface – UI | الأجزاء المرئية من التطبيق التي يتفاعل معها المستخدم، مثل الأزرار، والقوائم، والنوافذ. |
| إدارة الإصدارات | Version Control | نظام تتبع التغيرات والتحكم في إصدارات التطبيق المختلفة؛ لتسهيل عملية التحديث والصيانة. |
| لوحة المفاتيح الافتراضية | Virtual Keyboard | لوحة مفاتيح على الشاشة يمكن للمستخدمين النقر عليها لإدخال النصوص، بدلا من استخدام لوحة المفاتيح الفعلية. |

# الفصل الأول

## وثيقة معايير الحوسبة اللغوية العربية

- مقدمة
- وثائق معايير الحوسبة اللغوية العربية
- الأهمية والمسوغات
- تحديات معايير حوسبة اللغة العربية
- منهجية بناء معايير الحوسبة اللغوية العربية وتطويرها
- جهات تطوير معايير الحوسبة اللغوية
- معايير بناء المعاجم الإلكترونية العامة والمختصة
- معايير بناء منصات تعليم اللغة العربية لغةً ثانيةً
- معايير بناء وتوسيم المدونات اللغوية
- معايير دعم اللغة العربية في التطبيقات الحاسوبية
- المراجع



## مقدمة

تتعاظم أهمية إعداد الأطر النظرية والمعايير المرجعية في فروع الحوسبة التي تعالج البيانات اللغوية، ولا سيما في ظل التضخم غير المسبوق في حجم البيانات مع التقدم المذهل في القدرات الحاسوبية على معالجتها. وقد أسهم في انفجار إنتاج البيانات اللغوية ذلك الجيلُ الجديد لتطبيقات الإنترنت المعتمدة على المحتوى الذي ينشئه المستخدمون -بسهولة وفورية-، إلى جانب الأنظمة التوليدية للذكاء الاصطناعي التي بدأ يتزايد الاعتماد عليها في صناعة المحتوى اللغوي. وصاحبت هذه التطورات تحديات عديدة في ضبط جودة البيانات اللغوية، يترتب عليها مدى دقة المعالجات الحاسوبية المختلفة لها.



وتسـهم المعايـير والأطـر النظريـة لجمـع البيانـات اللغويـة وتخزينهـا ومعالجتهـا ونشـرها -إسـهاماً رئيسـاً- في مواجهـة مثـل هـذه التحديـات، وتسـاعد علـى تحسـين نتـائج المعالجـة الحاسـوبية للبيانـات اللغويـة في تطبيقـات متعـددة، بتوضيحهـا للطـرق والمعايـير المثلـى في التعامـل مـع البيانـات، ورفـع جـودة مخرجـات التطبيقـات اللغويـة الحاسـوبية، وتحقيـق الإفـادة المثلـى منهـا في سـياقات متعـددة، ولا سـيما فيمـا يتعلـق بمهمـات المعالجـة الحاسـوبية الخاصـة باللغـة العربيـة الـتي يتجـاوز عـدد المتحدثـين بهـا مئـات الملايـين في عالمنـا اليـوم، الـتي لا تـزال بأمـس الحاجـة إلى تكثيـف الجهـود وتشـجيع التعـاون البنـاء؛ لتطويـر معايـير ووثائـق مرجعيـة تخضـع للتحديـث المسـتمر في مختلـف مجـالات حوسـبة اللغـة؛ وذلـك لتحسـين نتـائج معالجـة البيانـات اللغويـة العربيـة المتزايـدة كمًّـا وكيفًـا في قوالـب ومنصـات وتطبيقـات حاسـوبية متعـددة.

وتقـدم وثائـق معايـير الحوسـبة اللغويـة العربيـة -الـتي أُعـدت بدعـمٍ وإشـرافٍ مـن مجمـع الملـك سـلمان العالـمي للغـة العربيـة- فصـولاً تشـمل الفـروع الرئيسـة للحوسـبة اللغويـة؛ لتجسـير الفجـوة في هـذا المجـال، والإسـهام في دفـع عجلـة الإبـداع والابتـكار في التطبيقـات الحاسـوبية المتعـددة للغـة العربيـة، وصقلهـا ورفـع فاعليتهـا وكفاءتهـا وجودتهـا.

ويتزامـن هـذا العمـل مـع زيـادة الاهتمـام بتحسـين جـودة البيانـات اللغويـة، وأهميـة حمايتهـا ومراعـاة خصوصيتهـا، وتطبيـق أخلاقيـات التعامـل الأمثـل معهـا؛ لأنهـا تشـكل البنيـة التحتيـة الـتي تعتمـد عليهـا تقنيـات الـذكاء الاصطناعـي التوليـدي والنمـاذج اللغويـة الضخمـة الـتي أحدثـت نقلـة نوعيـة وتغـيّرًا جذريًّـا في منهجيـة معالجـة مهمـات حوسـبة اللغـة التقليديـة، ووسـعت نطـاق اسـتعمالها في تطبيقـات وسـياقات متعـددة.

ويُعـد هـذا العمـل جـزءًا مـن مشـروع أكـبر يهـدف إلى تطويـر عـدد مـن الوثائـق الرئيسـة للمعايـير اللغويـة في عـدد مـن المجـالات الرئيسـة الـتي تسـتعمل فيهـا اللغـة العربيـة عامّـة. وهـو حلقـة في سلسـلة ممتـدة مـن جهـود المجمـع الدؤوبـة ومشـروعاته الطموحـة؛ لتعزيـز حضـور العربيـة ومكانتهـا، وتأكيـد فاعليتهـا، وترسـيخ اسـتخدامها في السـياقات الإقليميـة والمحليـة والعالميـة.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

وعلى الرغم من التقدم المتسارع في تقنيات معالجة اللغات البشرية؛ فإن المشكلات المعتادة في المعالجة الحاسوبية للغة لا تزال تشكل تحديًا مستمرًا للمختصين في اللسانيات الحاسوبية والمعالجة الآلية للغة، فعلى سبيل المثال لا تزال النماذج المتقدمة للذكاء الاصطناعي التوليدي تواجه عددًا من التحديات المتمثلة في انتهاك بعض مخرجاتها لحقوق الملكية الفكرية، وتحيز بعض نتائجها أو عدم دقتها؛ بسبب ضعف جودة تمثيل البيانات المستعملة في تدريب النموذج أو عدم صحتها، إضافة إلى تحدي تمييز المحتوى الزائف أو المُوجَّه، الذي يؤثر تأثيراً سلبياً على عمل ومخرجات نماذج الذكاء الاصطناعي التوليدي المعتمد على النماذج اللغوية الضخمة. إضافة إلى التحديات التقليدية للمعالجة الحاسوبية المرتبطة بمدى توافر البنية التحتية الرقمية، وتكلفة القدرات الحاسوبية لمعالجة البيانات الضخمة وتدريب النماذج اللغوية. ويأتي هذا العمل ليسهم في مواجهة مثل هذا التحديات، ويحسن نتائج تقييم وتوجيه مخرجات النماذج الحاسوبية للغة العربية، عبر وثائق معايير ترفع جودة المعالجة الحاسوبية لبيانات اللغة العربية في ستة تطبيقات رئيسة، تتضمن: (معايير بناء منصات تعليم اللغة العربية، ومعايير بناء المعاجم العامة، ومعايير بناء المعاجم المختصة، ومعايير بناء المدونات اللغوية العربية، ومعايير توسيمها، ومعايير دعم اللغة العربية في التطبيقات الحاسوبية).

ويقدم هذا الفصل إطارًا نظريًا لتطوير وثائق معايير الحوسبة اللغوية العربية، ويتضمن تمهيدًا لموضوعات معايير الحوسبة، ثم يتناول أهم مسوغات تطوير وثائق المعايير، وأبرز التحديات التي يواجهها الباحثون في هذه المجالات، ويقدم بعد ذلك شرحًا تفصيليًا للمنهجية المعتمدة في تطوير وثائق المعايير في هذا العمل. كما يستعرض الفصل أهم الجهات المهتمة بتطوير معايير الحوسبة اللغوية العربية خاصةً، والمعايير التقنية عامةً، ويُختم الفصل بمراجعة مختصرة لأهم الدراسات المعتمدة في وثائق معايير الحوسبة اللغوية العربية.

## وثائق معايير الحوسبة اللغوية العربية:

يقصد بوثيقة المعايير في سياق هذا العمل: دراسة موجزة تتضمن مجموعة من المعايير الموضوعية العامة والفرعية مع وصف مختصر لها، وتمثل قوائم المعايير في الوثائق المختلفة خلاصة تحليل عدد من البحوث والأنشطة التي قام بها فريق العمل؛ للتعرف على الجهود السابقة في كل مجال، ثم تلخيص أهم المعايير التي ترفع من جودة المعالجة الحاسوبية للبيانات والتطبيقات اللغوية التي تضمنتها الوثائق. وتسهم مراعاة هذه المعايير في تحسين المعالجة الحاسوبية لبيانات اللغة العربية، ورفع جودة مخرجات التطبيقات اللغوية.

وتتضمن معايير الحوسبة اللغوية مجموعة من التوصيات والمعايير والإرشادات التي ينبغي أخذها بعين الاعتبار في الموضوعات الستة التالية لوثائق المعايير، ويُمكن أن تَعتمد الجهات ذات العلاقة هذه المعايير؛ لتنظيم وتحسين مخرجات التطبيقات الحاسوبية الخاصة بمعالجة البيانات اللغوية العربية، ورفع جودة الحوسبة اللغوية. وقد روعي في كتابة وثائق المعايير أن تكون مهيَّأة للنشر على نحو مستقل، والالتزام بهذا المبدأ، يُسوِّغ وجود بعض التكرار أحيانًا في محتوى الوثائق؛ لعدم استعمال الإحالات المرجعية بينها.

وتشمل الوثائق أهم الموضوعات الرئيسة في حوسبة اللغة العربية، فيقدّم الفصل الأول إطارًا لوثائق معايير حوسبة اللغة العربية، ويحدد الثاني مجموعة من المعايير التي ينبغي مراعاتها في تصميم وتطوير وتنفيذ المنصات الإلكترونية الخاصة بتعليم اللغة العربية لغة ثانية، ويعرض الثالث مجموعة من المعايير الخاصة ببناء المعاجم الإلكترونية العامة، ويتناول الرابع معايير بناء وتطوير المعاجم المختصة أو معاجم المصطلحات، ويقدم الفصلان الخامس والسادس المعايير الخاصة ببناء المدونات وتوسيمها وتقويمها، ويتضمن الفصل الأخير مجموعة من المعايير الخاصة بدعم اللغة العربية في التطبيقات والمنصات الحاسوبية.

ويشتمل كل فصل على ملخص لمحتواه، ثم مقدمة وتمهيد تتضمن التعريف بالمجال وأهميته ومدخلا تمهيديا لمجال المعايير، ثم استعراض ومراجعة لأهم الدراسات السابقة مع تركيز فيما يتصل بموضوع الوثيقة، ووصف مختصر لنطاق المعايير وحدودها، ثم يُختم الفصل بسرد مختصر للمعايير الخاصة بالمجال، مصنفة في فئات متعددة بحسب محتواها. ومن المهم تأكيد ضرورة التحديث والتقويم والمراجعة المستمرة لوثائق المعايير بمختلف مجالاتها واختصاصاتها، ولا سيما في المعايير المرتبطة بالتقنية؛ للتطور السريع للتقنيات الحاسوبية المختلفة لمعالجة اللغة، ولضمان مواكبتها لأحدث تلك التقنيات، ومطابقتها لأهم المعايير والمواصفات والمقاييس الفنية المختصة المحلية والعالمية.

## الأهمية والمسوغات:

لا تزال المعايير والوثائق المرجعية هي الدافع الرئيس للتقدم والتطور الذي شهدته معظم التقنيات المختلفة في العقود الأخيرة؛ لأنها تُشكل اللغة المشتركة التي يتواصل بها المختصون في المجال التقني، فهي تسهم في رفع مستوى جودة المشروعات التقنية، وتقود التقدم والابتكار في مجالات الحاسوب والاتصالات والمعلومات خاصة، كما أن تبنيها في المشروعات التقنية المختلفة يعزز من زيادة مستوى التواصل الفعال، وكفاءته، ويسهم في التعاون المثمر بين الجهات والأفراد المشاركين في إنجاز المهمات المختلفة. ولذلك كانت وثائق المعايير والمواصفات والمقاييس خطوة مبدئية وضرورية، أسهمت في توسيع نطاق استعمال كثير من التقنيات الحديثة، ورفعت من مستوى جودتها. وقد يرى البعض أن مفهوم المعايير يناقض منهجية الإبداع والابتكار والتجديد التي من أهمها التحرر من القيود السابقة وإضافة الجديد، ولكن عند تأمل العلاقة بين المعايير وأنواع الابتكار المختلفة في المجال التقني، نجد أن العلاقة بينهما علاقة تبادلية وتكاملية؛ فقد أسهم الالتزام بالمعايير التقنية على مدى السنوات الماضية في توسيع نطاق استعمال كثير من التقنيات (مثل: تقنية البلوتوث[1] - وتقنيات الجيل الخامس للاتصالات اللاسلكية)،

(١) استعملت تقنية البلوتوث في بداية ابتكارها طريقة للتوصيل غير السلكي المحدود بين الحاسوب والطابعة، وانتقلت بعد اعتمادها على المعايير والمحددات التقنية العالمية إلى حالات استخدام متعددة وواسعة في التوصيل بين الأجهزة (للمزيد، انظر: Šimunić & Pavić, ٢٠٢٠).

وقـاد ذلـك إلى ابتـكار عـدد مـن التقنيـات الحديثـة، جـاءت اسـتجابة لاحتياجـات جديـدة ومختلفـة لـم تكـن متوافـرة في السـياقات الأوليـة المحـدودة لهـذه التقنيـات؛ ممـا عـزز القيمـة المضافـة لاسـتعمال مثـل هـذه التقنيـات في البيئـات الجديـدة. ومـن ثَـمّ يمكـن القـول إن إعـداد المعايـير والالـتزام بتطبيقهـا أسـاسٌ لابتـكارات تقنيـة عديـدة في مجـالات متنوعـة، وعلـى رأسـها مجـال الحوسـبة اللغويـة بـكل تطبيقاتـه وفروعـه.

وقـد طُـورت هـذه الوثائـق بنـاء علـى عـدد مـن الدوافـع والمسـوغات الـتي تهـدف إلى تحسـين نتـائج حوسـبة اللغـة العربيـة في تطبيقـات وسـياقات متعـددة، وسـيُوضّح بالتفصيـل في كل وثيقـة علـى حـدة، ويمكـن تلخيـص أهـم المسـوغات والدوافـع العامـة في النقـاط الآتيـة:

- تأكيـد أهميـة العنايـة بتحضـير البيانـات اللغويـة، وتوحيـد طـرق معالجتهـا وتصفيتهـا مـن البيانـات غـيرذات الصلـة الـتي قـد تؤثـر في جـودة نتـائج المعالجـة الحاسـوبية.
- تقديـم مجموعـة مـن المعايـير لجمـع البيانـات اللغويـة وتوسـيمها وعرضهـا في قوالـب إلكترونيـة متعـددة؛ لضمـان الاتسـاق والانتظـام في عمليـات المعالجـة الحاسـوبية في مراحـل التحليـل اللغـوي المختلفـة، بدايـة مـن المسـتوى الكتـابي الصـوتي حـتى الوصـول إلى المسـتويات الدلاليـة والتداوليـة المتقدمـة.
- ضمـان التمثيـل الشـامل والمتـوازن لكافـة المسـتويات والتباينـات اللغويـة، وتجنـب التحـيز المقصـود وغـير المقصـود في تمثيـل البيانـات اللغويـة الـذي يؤثـر تأثـيرا سـلبياً في صحـة ودقـة مخرجـات النمـاذج اللغويـة الحاسـوبية.
- أهميـة توظيـف آليـات لمراقبـة جـودة البيانـات اللغويـة وتمييـز المشـكلات والفجـوات فوريـاً، ومحاولـة تحييدهـا أو التقليـل مـن تأثيرهـا في نتـائج التطبيقـات اللغويـة المتعـددة.

- الحرص على مراعاة ما يناسب من المعايير الدولية ذات الصلة وأهمية التحديث المستمر للمعايير؛ لتواكب التقنيات الحديثة المتسارعة، مع التأكيد على أهمية تمثيلها في الوقت نفسه للاستعمال اللغوي المستهدف بكل مستجداته.

- تقدم المعايير أطرًا نظرية تمكن من بناء المصادر اللغوية المحوسبة مثل المعاجم، والمدونات اللغوية، وتساعد على ضمان اتساق هذه المصادر وتطبيقها لأعلى مواصفات الجودة في تصميم وتقييم المصادر اللغوية المحوسبة.

- تسهل المعايير من نقل وتبادل ومشاركة التطبيقات والموارد في اللغة نفسها أو بين عدد من اللغات المختلفة، ولا سيما اللغات الأقل حظًّا في توافر المصادر اللغوية؛ إذ إن عند توحيد المعايير وأنظمة توسيم وتمثيل البيانات، يمَكّن من توظيف الترجمة لتجسير الفجوة في توافر البيانات بين اللغات غنية الموارد وغيرها من اللغات الأقل من حيث الموارد اللغوية.

- تمكن المعايير الموحدة من تطوير مصفوفات التقويم للمصادر والتطبيقات اللغوية، وإجراء المقارنات المرجعية بين التطبيقات والموارد اللُّغويَّة؛ التي تسهم في رفع الجودة وزيادة فعالية المقارنات المرجعية وأثرها، إضافة إلى تعزيز المنافسة وتسهيل تتبُّع التَّطورات في تطوير التطبيقات والموارد اللغوية.

- تساعد المعايير على ضمان جودة بناء المصادر اللُّغوية؛ عن طريق توضيح الممارسات والأطر المثلى لتطويرها وتقويمها، إضافة إلى مساعدتها في تَعَرُّف مواضع الضعف والخلل، وتتبع الفجوات اللغوية في نطاق الموارد الرقمية، وتقديم الأسس التي يعتمد عليها في تقويم التطبيقات والمصادر اللغوية، إضافة إلى إمكانية استعمالها في المجالات التعليمية والتدريبية.

- تساعد المعايير على ضمان اتساق عملية بناء الموارد والتطبيقات اللغوية وتقويمها، إضافة إلى تيسير إجراءات توسيعها وزيادة سعة نطاقها.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً



## تحديات معايير حوسبة اللغة العربية:

إعـداد المعايـير لـه أهميـة في مختلـف المجـالات، وتـزداد أهميتـه في معالجـة اللغـة العربيـة؛ لـثراء سـماتها الصرفيـة وطبيعـة خصائصهـا الاشـتقاقية، إضافـة إلى تحديـات التمثيـل الكتـابي للغـة العربيـة واختلافاتهـا الصوتيـة واللهجيـة (Habash, 2010; Farghaly & Shaalan 2009; Zribi et al. ,2014).

وتعتمـد جـودةُ كثـيرٍمـن مجـالات الحوسـبة اللغويـة علـى معايـيروأطـرٍ مرجعيـةٍ راسـخة، ولاسـيّما في توسـيم المدونـات، وأطـر التمثيـل الـدلالي، وبنـاء المـوارد اللغويـة وتطويرها، فضلًا عـن دور المعايـير في ضبط الجودة وتعزيز نتـائج التقويم والمقارنات المرجعية بين المـوارد والتطبيقـات. وقـد قدّمـت دراسـاتٌ مرجعيـة أطرًا ومعايـير في هـذا المجـال، ومنها: (Leech, 1993; Palmer & Xue, 2013; Pustejovsky & Stubbs, 2012; Nivre et al., 2020; Nivre et al., 2016; Ide & Pustejovsky, 2017; Bunt et al., 2010)

ومـن التحديـات الـتي تواجـه مطـوري المعايـيروالأطـر المرجعيـة في هـذا المجـال: التنـوع اللغـوي الـذي يُصَعِّـب توحيـد المعايـير في ظـل الاختلافـات اللغويـة واللهجيـة وتعقيداتهـا المختلفـة، إضافـة إلى نـدرة البيانـات اللغويـة الرقميـة الـتي يمكـن الوصـول إليهـا في اللغـات أو اللهجـات قليلـة الاسـتعمال، وكذلـك الوقـت والجهـد اللـذان يتطلبهمـا تطويـر مثـل هـذه الوثائـق، وتحديثهـا ومراجعتهـا المسـتمرة. ولذلـك اتجـه الاهتمـام إلى تطويـر النمـاذج اللغويـة الحاسـوبية الضخمـة المعتمـدة علـى النمـاذج التحويليـة، والمدربـة علـى بيانـات لغويـة ضخمـة غـيرموسـومة أو خاضعـة لمعايـيرمحـددة. وعليـه، يجـب علـى مطـوري المعايـيروالأطـر المرجعيـة لحوسـبة اللغـة مواكبـة هـذه التطـورات وتقديـم طـرق مناسـبة وسـريعة لأتمتـة إعـداد وتطبيـق المعايـيراللغويـة في مهمـات المعالجـة الحاسـوبية.

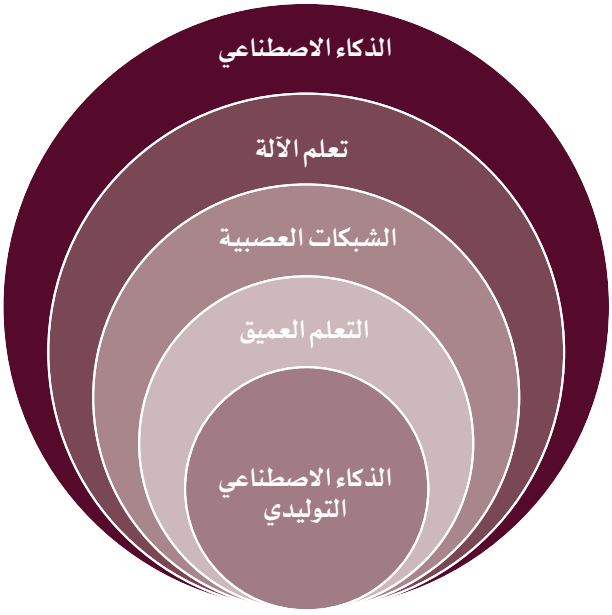
وتفـرض الاتجاهـات الحديثـة في مجـال معايـيرحوسـبة اللغـة، أن تواكـب المعايـير اللغويـة في معالجـة اللغـة الطبيعيـة (NLP) التقـدم الحـادث في معالجـة البيانـات متعـددة الوسـائط واللغـات. ومـع تزايـد اعتمـاد أنظمـة معالجـة اللغـة الطبيعيـة علـى دمـج البيانـات

اللغويـة مـع عـدد مـن أنـواع البيانـات الأخـرى -مثـل الصـور والمـواد متعـددة الوسـائط- يصبـح مـن الضـروري تطويـر معايـير جديـدة لتنسـيق التوسـيم اللغـوي عـبر الوسـائط المختلفـة. إضافـة إلى توسـيع المعايـير متعـددة اللغـات الحاليـة لدعـم اللغـات منخفضـة المـوارد والمهـددة بالاندثـار لتعزيـز التنـوع اللغـوي في البيانـات، ويشـمل ذلـك معالجـة الميزات اللغوية الفريـدة، مثـل: اللهجـات النـادرة، وظواهـر التغـير اللغـوي؛ لضمـان التمثيـل الحاسـوبي الـكافي للغـات قليلـة الاسـتعمال.

ومـع تطـور نمـاذج التعلـم العميـق والنمـاذج العصبيـة والنمـاذج اللغويـة الكبـيرة (LLMs)، يلـزم أن تتكيـف المعايـير اللغويـة مـع هـذه المناهـج الجديـدة للمعالجـة الحاسـوبية، وأن تتضمـن موضوعـات تقييـم عمليـات المحـولات الرقميـة للغـة، مثـل: تضمـين الكلمـات وتمثيـل معانيهـا (embeddings)، وعمليـات تخصيـص المحتـوى اللغـوي (fine-tuning)، وآليـات التفسـير الضروريـة؛ لضمـان الاتسـاق والجـودة في بيئـة النمـاذج الحاسـوبية العصبيـة. إضافـة إلى مواكبـة تحـول التوسـيم اللغـوي نحـو الأسـاليب شـبه المؤتمتـة أو المدعومـة بالـذكاء الاصطناعـي، وإعـداد أطـر إرشـادية جديـدة للحفـاظ علـى الجـودة والاتسـاق، مـع الإفـادة في الوقـت نفسـه مـن نتـائج التوسـيم الآلي. كمـا ينبغـي تطويـر مقاييـس لتقييـم نمـاذج الـذكاء الاصطناعـي التوليـدي مـن حيـث دقتهـا اللغويـة، ومسـتوى التحـيز فيهـا، ومـدى مراعاتهـا للحقـوق الفكريـة والاعتبـارات والمبـادئ الأخلاقيـة.

وتتضمـن الاتجاهـات المسـتقبلية ضـرورة توسـيع عمليـات التوسـيم اللغـوي وتحسـين معايـير التقييـم؛ إذ يمكـن أن تسـرّع عمليـات التوسـيم المدعومـة بالـذكاء الاصطناعـي والمصـادر الجماعيـة (crowdsourcing) مـن بنـاء المـوارد اللغويـة وتطويرهـا، مـع مراعـاة المعايـير وآليـات مراقبـة الجـودة لضمـان التوافـق مـع الأطـر الموحـدة. كمـا تـبرز أهميـة تطويـر معايـير تقييـم متغـيرة وديناميكيـة ومواكبـة للتقـدم في مناهـج معالجـة مهمـات اللغـة الطبيعيـة؛ لتصبـح أكـثر توافقـاً مـع التطبيقـات المتنوعـة لمعالجـة اللغـة الطبيعيـة، مثـل: التلخيـص الآلي، وأنظمـة المحادثـة المعتمـدة علـى الـذكاء الاصطناعـي.

ومـن الأسـس المهمـة الـتي ينبغـي الاعتمـاد عليهـا في هـذا المجـال: تعزيـز التعـاون بـين التخصصـات البينيـة المختلفـة، ويظهـر (الشـكل ١) عـددًا مـن التخصصـات المتصلـة بالـذكاء الاصطناعـي ومعالجـة اللغـات وفروعهـا المختلفـة الـتي ينبغـي أخذهـا بعـين الاعتبـار.

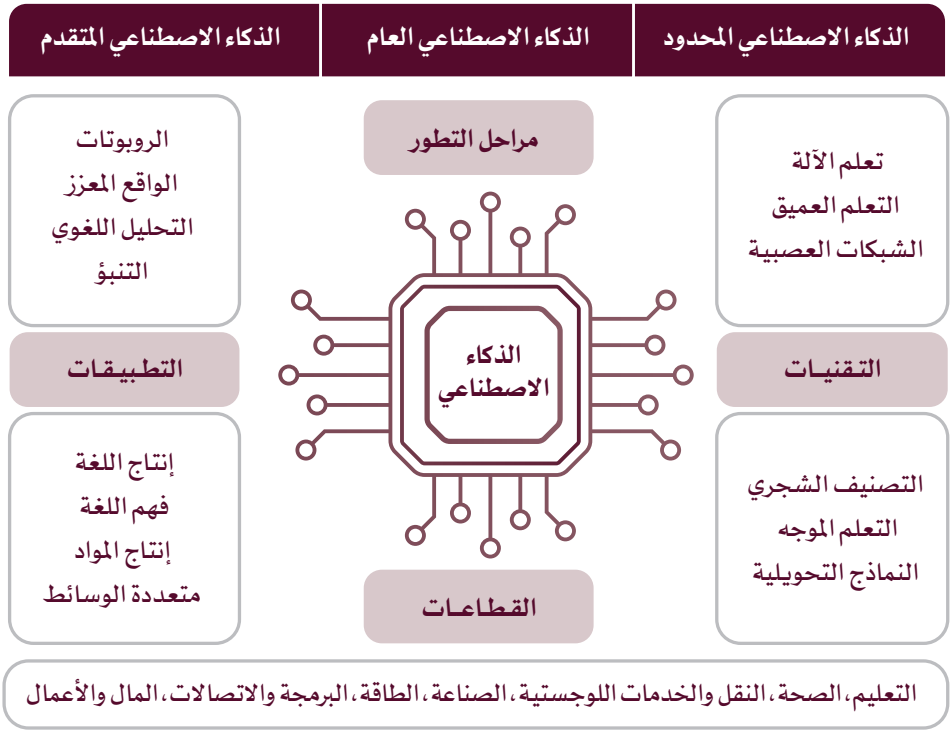



الشكل ١: أهم التخصصات الفرعية لنماذج الذكاء الاصطناعي.

ويتطلـب التطـور في نطـاق تطبيقـات معالجـة اللغـة والـذكاء الاصطناعـي، تطويـر فئـات متعـددة للمعايـير تتضمـن المجـالات المسـتجدة وتطبيقاتهـا المختلفـة والجوانـب الأخلاقيـة لاسـتعمالها.

كما يجب أن ينعكس التوسع غير المسبوق في تطبيقات الذكاء الاصطناعي التوليدي على المعايير والأطر المنظمة للتقنيات والتطبيقات المرتبطة بها. ويظهر (الشكل ٢) مراحل تطور الذكاء الاصطناعي التوليدي التي نشهد بدايتها في قدرات تطبيقات الذكاء الاصطناعي التوليدي محدود المجال، التي يتوقع أن تتبعها مرحلة الذكاء الاصطناعي

العـام والمتقـدم، كمـا يظهـر أمثلـة لعـدد مـن تقنيـات الـذكاء الاصطناعـي وتطبيقاتهـا في قطاعـات مختلفـة.




الشكل ٢: مراحل تطور الذكاء الاصطناعي التوليدي وأهم تقنياته وتطبيقاته.

ويمكـن أن تقـدم المجـالات اللسـانية المعرفيـة ونظريـات التفاعـل بـين الإنسـان والحاسـوب رؤى جديـدة تسـاعد على تطويـر أطـر مرجعية تُبـرز دور الـذكاء البشـري في تطوير نمـاذج الـذكاء الاصطناعـي. كمـا أن المعايـير مفتوحـة المصـدر ومشـاركة البيانـات، تُعَـدّ مـن العوامـل الأساسـية لتعزيـز التعـاون العالـمي، وزيـادة نسـبة وحجـم المـوارد اللغويـة المتاحـة والقابلـة للتكيـف في سـياقات وتطبيقـات متعـددة. ومـن خـلال معالجـة هـذه الأولويـات، يمكـن للمعايـير اللغويـة في المسـتقبل دفـع الابتـكار ودعـم التنـوع اللغـوي وتعزيـز الممارسـات الأخلاقيـة والتمثيـل العـادل في تطويـر تقنيـات معالجـة اللغـة البشـرية وتقييمهـا.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً



## منهجية بناء معايير الحوسبة اللغوية العربية وتطويرها:

يعتمد تطوير الوثائق الحالية الخاصة بمجالات حوسبة اللغة العربية منهجيةً دورية تتكون من مراحل مترابطة ومتكاملة، يوضحها (الشكل ٣)، **وذلك على النحو التالي:**

١- مرحلة تحديد مجال الوثيقة وسياقها وأهدافها.

٢- مرحلة جمع البيانات والمراجع، ولا سيما ما يتعلق بالمعايير السابقة ذات الصلة بموضوع الوثيقة.

٣- دراستها وتحليل بياناتها ومقارنتها بموضوع وأهداف الوثيقة المطورة؛ لاستخلاص ما يناسب من المعايير والمؤشرات وتبنيها واستبعاد ما ليس له علاقة من المعايير أو المراجع السابقة.

٤- مرحلة تحليل واقع اللغة العربية مقارنة باللغات الأخرى في الوثائق السابقة، وتحديد أهم المعايير التي يمكن أن تتضمنها وثائق حوسبة اللغة العربية.

٥- كتابة المسودة الأولى للمعايير.

٦- مرحلة المراجعة والتقييم، التي يقوم بها الفريق العلمي لإضافة أو حذف ما يناسب من المعايير بناء على ذلك.

٧- مرحلة التحكيم.

٨- وأخيراً: مرحلة اعتماد الوثائق ونشرها.

وتشير العلاقة الدائرية التي تظهر بين مراحل تطوير وثائق الحوسبة اللغوية في (الشكل ٣) إلى أنها عملية تكاملية مستمرة، يجب أن تخضع للتحديث والتطوير المستمر لتوافق أحدث التقنيات والإجراءات المتبعة عالميًّا في مجالاتها المختلفة.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

ومـن الخطـوات المهمـة الـتي مـرت بهـا الوثائـق - قبـل البـدء في تنفيـذ هـذه المراحـل وتحديـد مجالاتهـا وأهدافهـا - دراسـة وتحليـل الاحتيـاج المتوقـع لوثائـق المعايـير عـن طريـق التواصـل مـع جميـع الأطـراف ذات العلاقـة، وجمـع البيانـات المرتبطـة بتحديـد الاحتياجـات، مثـل تحليـل الفجـوة في التطبيقـات الحاليـة، والتعـرف علـى التحديـات اللغويـة والتقنيـة الـتي يمكـن أن تسـهم هـذه المعايـير في تجاوزهـا. وتهـدف مرحلـة البحـث وجمـع البيانـات إلى استكشـاف الدراسـات ذات العلاقـة لفهـم المناهـج الملائمـة لتطويـر الوثائـق، واستكشـاف الأطـر والوثائـق المتاحـة؛ للتأكـد مـن الاتسـاق وتجنـب التكـرار أو الاختـلاف في محتـوى الوثائـق.




الشكل ٣: مراحل إعداد وثائق معايير الحوسبة اللغوية

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً



وقـد روعـي في منهـج إعـداد الوثائـق الأسـس المتبعـة في بنـاء وتطويـر وثائـق المعايـير في الجهـات الأكاديميـة والمنظمـات الدوليـة المختصـة، مـع مراعـاة السـياق والإمكانيـات المتاحـة في أثنـاء تطويـر الوثائـق الحاليـة. وتبـدأ دورة تطويـر المعايـير الدوليـة عـادة بمرحلـة الإعـداد والتحضـير وتكليـف اللجـان المختصـة، ثـم صـدور المقـترح الأولي الـذي يخضـع للمراجعـة والتحكيـم، ثـم التطويـر والاعتمـاد والنشـر. فعلـى سـبيل المثـال، تمـر وثيقـة المعايـير العالميـة في (منظمـة الآيـزو) بسـت مراحـل؛ تبـدأ مـن المرحلـة الأوليـة لتحديـد مجـال الوثيقـة وأهدافهـا، ثـم المرحلـة التحضيريـة الـتي تليهـا بدايـة التنفيـذ وإعـداد المعايـير ونشـر النسـخة الأولى داخليًـا، ثـم تحكيمهـا لـدى الخـبراء والمختصـين، وإقرارهـا ونشـرها خارجيًـا مـن قبـل الأمانـة العامـة للمنظمـة.

وتمثـل وثائـق المعايـير موضوعـات رئيسـة في الحوسـبة اللغويـة، اختـيرت بنـاءً علـى تحليـل الاحتيـاج المتوقـع لرفـع جـودة التطبيقـات اللغويـة. وتراعـي هـذه المعايـير خصائـص العربيـة في مسـتويات التحليـل المختلفـة، ممـا يسـهم في تحسـين كفـاءة معالجـة البيانـات ودقـة مخرجـات النمـاذج الحاسـوبية. كمـا أن جـودة البيانـات اللغويـة تعـد عامـلًا حاسـمًا في أداء النمـاذج وصحـة نتائجهـا، إذ تضمـن تمثيـلًا أفضـل للسـياقات المسـتهدفة وتقلـل مـن التحيـز والأخطـاء. ويسـاعد ذلـك علـى تعميـم نتـائج المعالجـة وتلبيـة احتياجـات مسـتخدمي التطبيقـات اللغويـة المتنوعـة.

وقـد مثـل التطـور الحديـث في تقنيـات ونمـاذج الـذكاء الاصطناعـي التوليـدي تحـولاً جذريًـا في منهجيـات المعالجـة الآليـة للغـة البشـرية؛ إذ تعتمـد النمـاذج اللغويـة الكبـيرة علـى بيانـات لغويـة ضخمـة، وتحقـق مسـتويات غـير مسـبوقة مـن الدقـة في فهـم اللغـة وإنتاجهـا مقارنـة بالنمـاذج الإحصائيـة أو القاعديـة التقليديـة.

وأسـهم تبـني تقنيـات الـذكاء الاصطناعـي التوليـدي في تحسـين مخرجـات معظـم مهمـات معالجـة اللغـة الطبيعيـة، وتوسـيع مجـالات الإفـادة منهـا -كمـا يوضـح (الشـكل٢)- إذ

تتضمن على سبيل المثال: صناعة المحتوى اللغوي الإبداعي، ومهمات التحليل اللغوي بكافة مستوياتها، وتحسين نتائج تطبيقات الترجمة والتلخيص وإعادة الصياغة اللغوية، وتخصيص المحتوى في التجارب الرقمية، وغيرها من الوظائف والتطبيقات التي كانت مقتصرة في السابق على العنصر البشري؛ لاحتياجها إلى درجة عالية من الذكاء ومهارات حل المشكلات المعقدة.

## جهات تطوير معايير الحوسبة اللغوية:

تنشط منظماتٌ دولية عدّة - ذات صلة وثيقة بالمجال التقني- في إعداد المعايير ونشرها؛ نظرًا إلى أهمية تلك المعايير في ضبط الجودة ورفعها لمستوى تطوير البرمجيات والمشروعات التقنية. ومن أهم المنظمات العالمية المختصة في إعداد ونشر المعايير التقنية: منظمة الآيزو العالمية للمعايير (International Organization for Standardization – ISO)، والاتحاد الدولي للاتصالات (International Telecommunication Union – ITU)، واللجنة العالمية الكهروتقنية (International Electrotechnical Commission – IEC)، ومعهد مهندسي الكهرباء والإلكترونيات (Institute of Electrical and Electronics Engineers – IEEE). وقد بدأت بعض هذه المنظمات العمل في مجال إعداد ونشر المعايير التقنية منذ أكثر من مئة عام، ولا تزال تعمل في المجال نفسه حتى وقتنا المعاصر[٢]، وتركز أغلب المنظمات الدولية الخاصة بالمعايير -في جُل أعمالها- في إعداد الوثائق المرجعية والمعايير الدولية الخاصة بالتطبيقات التقنية المختلفة، وتحتل غالبًا المعايير التقنية النسبة الأكبر في سجلات وثائق المعايير والمواصفات التي تصدرها وتحدثها هذه المنظمات باستمرار.

(٢) مواقع المنظمات المذكورة :// IEEE www.ieee.org ITU www.itu.int. IEC www.iec.ch ISO https: www.iso.org

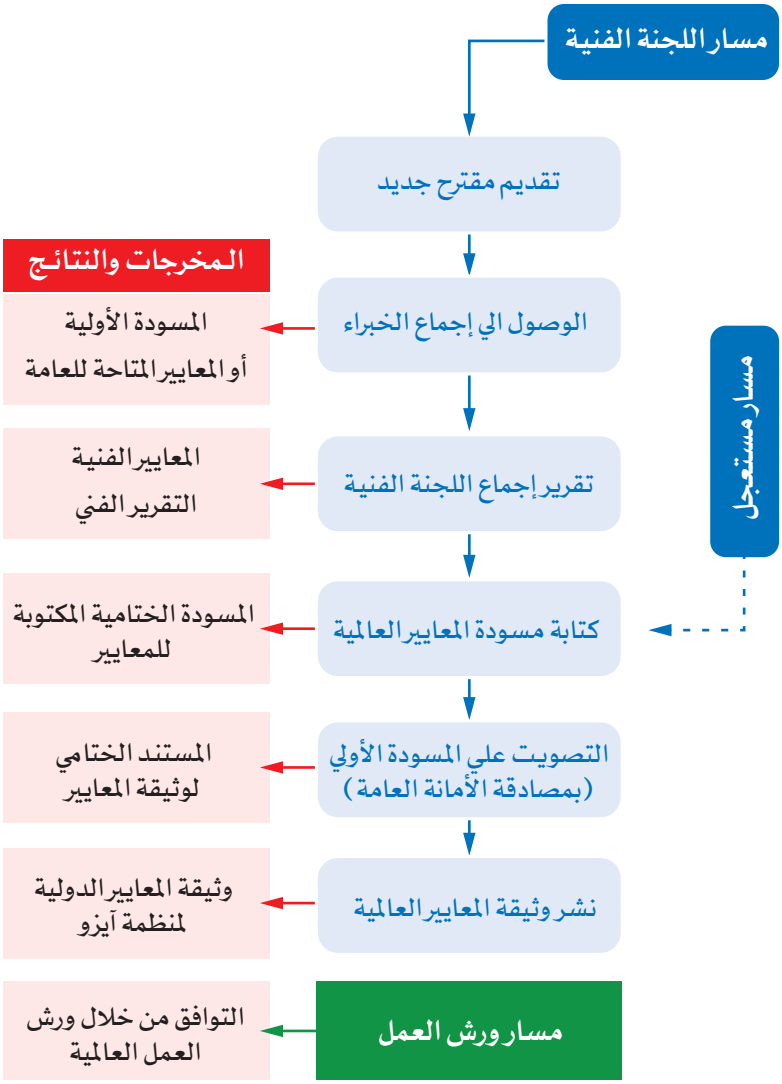
ونقدم فيما يلي نبذة مختصرة عن أهم الجهات العالمية التي تُعنى بإصدار وثائق المعايير في التطبيقات اللغوية والتقنية؛ وذلك لاعتماد عدد من معاييرها في الوثائق المتصلة بمحاور وثائق معايير الحوسبة اللغوية.

**منظمة الآيزو العالمية للمعايير (International Organization for Standardization – ISO):**

تُعَد من أهم المنظمات العالمية في إعداد وثائق المواصفات والمعايير على مستوى العالم، وهي منظمة غير حكومية، تضم في عضويتها ممثلين من أكثر من ١٦٤ دولة، وتشكل جمعيتها العامة -التي تعقد اجتماعاتها سنوياً- أعلى سلطة تنفيذية، فيما تتولى أعمالها الأمانة العامة من مقرها في جنيف بسويسرا. وقد تأسست المنظمة في لندن عام ١٩٤٧؛ لتنسيق الجهود الدولية، والتعاون في توثيق وإدارة وإنتاج المعايير الدولية لأغراض مختلفة.

وتضم المنظمة أكثر من ٧٨٥ لجنة فنية للمعايير، ونشرت أكثر من ٢٢٦٣١ وثيقة في المجالات التقنية المتنوعة وما يتصل بها، تشكّل المعايير التقنية منها نحو ٢١,٧٪. وتسهم المنظمة في تطوير معايير تدعم أهداف التنمية المستدامة التي تشرف عليها منظمة الأمم المتحدة.

وتمر صياغة وثائق المعايير في المنظمة بمجموعة من الخطوات والمراحل الرئيسة تبدأ بدراسة مقترح الوثيقة، ثم صناعة إجماع الخبراء والمختصين حولها، ثم نقل النتائج إلى اللجنة الفنية المختصة واتخاذ قرارات مُجْمَع عليها بخصوص بنود وثيقة المعايير، وبعد ذلك تُكتب المسودة الأولى للمعايير، ثم تخضع لمراجعة الأمانة العامة التي تعتمد بعد ذلك القرار بنشرها. ويوضح (الشكل ٤) هذه المراحل والمهمات التي تجرى فيها بالتفصيل.




الشكل ٤: مسار تطوير وثيقة المعايير في منظمة آيزو للمعايير العالمية[٣].

(٣) للمزيد، راجع المصدر: http://www.iso.org/iso/home/standards_development/deliverables-all.htm

**الاتحاد الدولي للاتصالات (International Telecommunication Union – ITU):**

يُعدّ من أقدم المنظمات العالمية المختصة في إعداد معايير وبروتوكولات الاتصالات السلكية واللاسلكية، وتسهيل استعمال تقنيات الاتصال عالمياً، عبر إعداد ونشر مجموعة من المعايير التقنية في مجالات الاتصالات وتقنية المعلومات. تأسس عام ١٨٦٥ في باريس باسم الاتحاد الدولي للتلغراف، وهو اليوم وكالة متخصصة تابعة لمنظمة الأمم المتحدة. وللاتحاد إضافة إلى الأمانة العامة ثلاثة قطاعات رئيسة: قطاع الاتصالات الراديوية، وقطاع تقييس الاتصالات، وقطاع تنمية الاتصالات. وتضم عضويته ١٩٣ دولة وأكثر من ألف شركة وجامعة.

**اللجنة العالمية للتقنيات الكهروتقنية (International Electrotechnical Commission – IEC):**

تأسست في عام ١٩٠٦، ومقرها في جنيف بسويسرا. وتعد من أقدم المنظمات الدولية المختصة في إعداد ونشر مجموعة من المعايير في مجال الكهرباء والإلكترونيات، وتضم وثائق المنظمة مجموعة واسعة من المجالات ذات العلاقة، مثل: تقنيات الألياف البصرية والبطاريات وأشباه الموصلات والنانو وغيرها. وتضم عضوية المنظمة ١٧٠ دولة، وقد نشرت المنظمة أكثر من ١٠ آلاف وثيقة للمعايير، وأكثر من مليون شهادة اعتماد، ويعمل في إعداد وثائقها أكثر من ٣٠ ألف خبير يمثلون خلفيات وخبرات متنوعة.

**معهد مهندسي الكهرباء والإلكترونيات**
**(Institute of Electrical and Electronics Engineers – IEEE):**

يُعد من أهم المنظمات العالمية التي تعنى بتطوير المعايير والأدلة المرجعية التقنية في مجالات الاتصالات وتقنية المعلومات والذكاء الاصطناعي، ونشرت المنظمة مجموعة من الوثائق الخاصة بأخلاقيات الذكاء الاصطناعي، وتعلم الآلة، ومعالجة البيانات واللغات الطبيعية.

وتقدم الأجزاء التالية في هذا القسم عرضًا موجزًا لأهم الوثائق والأطر والدراسات السابقة في محاور معايير الحوسبة اللغوية العربية التي تعالجها هذه الوثيقة.

## معايير بناء المعاجم الإلكترونية العامة والمختصة:

بدأ الاهتمام مبكرًا بإعداد المعايير والأطر المرجعية واعتمادها؛ لبناء وتقويم المعاجم والموارد اللغوية الإلكترونية؛ لأثرها البالغ في تعزيز جودة البيانات المعجمية، وتمكين مشاركة ودمج المصادر المعجمية محليًّا وعالميًّا في بيئات وسياقات مختلفة. وأسهمت منظمات إعداد ونشر المعايير الرقمية في نشر ثقافة الجودة والاتساق في الصناعة المعجمية الرقمية. وفي أدبيات المعاجم الإلكترونية نجد تصنيفات متعددة للمعاجم وفقًا لأغراض استعمالها المختلفة، ومن أهمها تقسيم المعاجم إلى عامة ومختصة أو معاجم اصطلاحية، إضافة إلى تقسيم المعاجم الإلكترونية إلى معاجم تستهدف استعمال الآلة وأخرى تعد للاستعمال البشري. وقد جذب كل نوع من هذه المعاجم مجموعة كبيرة من الباحثين لدراستها وإعداد المعايير والأطر النظرية لبنائها وتقويمها.

وقد زادت مؤخرًا أهمية هذه الوثائق بسبب التطور غير المسبوق في تدفق البيانات اللغوية من منصات صناعة المحتوى الرقمي، والحاجة العملية لتحليل هذه البيانات، وإثراء المعاجم اللغوية بها، والإفادة من نتائج تطبيقات المعالجة الآلية للغة في مجالات متعددة. ولهذا تهدف معظم وثائق معايير المعاجم العامة إلى محاولة توحيد عمليات ترميز البيانات المعجمية الرقمية وتمثيلها في نموذج حاسوبي مرن يراعي خصائص مجموعة من الأنظمة اللغوية المختلفة، ويمكن من خلال هذا النموذج تبادل البيانات المعجمية وتيسير طرق صناعتها.

ومن أهم الوثائق التي تمثل الجهود الدولية في هذا الصدد، الإطار المرجعي لتمثيل وترميز البيانات المعجمية (Lexical Markup Framework – LMF)، الذي يمثل تتويجًا لجهود سابقة من جهات متعددة، قدمت خلال العقدين الماضيين مجموعة من المقترحات

لنمـاذج حاسـوبية يمكـن أن تعتمـد علـى نطـاق واسـع لتمثيـل البيانـات المعجميـة، وتدعـم خاصيـة تعـدد اللغـات والتطبيقـات (Šimunić & Pavić, 2020). وتهـدف الوثيقـة في آخـر تحديـث لها (ISO 24613-1:2024) إلى تحسـين تمثيـل بيانـات المعاجـم الإلكترونيـة، والتكامـل الفعـال في تطبيقاتها، ودعمها لتقنيـات الترجمـة المختلفـة. ويقـدم النمـوذج هيـكلا موحـدا لتمثيـل البيانـات المعجميـة العامـة، ويمكـن الإفـادة منـه في إعـادة اسـتعمال المصـادر اللغويـة، وتحقيـق الاتسـاق والتوافـق بـين التطبيقـات اللغويـة المختلفـة، إضافـة إلى شـموله لـكل مسـتويات التمثيـل اللغـوي الصـوتي والصـرفي والنحـوي والـدلالي.

وقدمـت وثيقـة (ISO 1951:2007) إطـارًا لتقديـم وعـرض المعاجـم الإلكترونيـة، وتشـمل الوثيقـةُ المعاجـمَ العامـة والمختصـة والمعاجـم أحاديـة أو متعـددة اللغـات، ومـن أهـم أهدافهـا: توحيـد طريقـة عـرض المعلومـات في المعاجـم، وتحديـد الرمـوز المسـتخدمة في المداخـل المعجميـة إضافـة إلى تسـهيل تبـادل البيانـات المعجميـة بـين اللغـات المختلفـة. ويحـدد المعيـار نظـام ترميـزموحـد للمعلومـات اللغويـة يتضمـن رمـوزاً خاصـة للعناصـر النحويـة والصرفيـة، ويضـع قواعـد لتنظيـم المداخـل المعجميـة وترتيبهـا، كمـا يحـدد طـرق عـرض المعلومـات الصوتيـة والنطقيـة، ويقـدم إرشـادات لتمثيـل العلاقـات الدلاليـة. ومـن المميـزات الإضافيـة لهـذا المعيـار، تسـهيله للتحويـل والنقـل بـين الأنظمـة المختلفـة، ودعـم خصائـص البحـث المتقـدم والتصفيـة، وتمكـين الربـط بـين المداخـل المعجميـة، وتحديـث محتـوى البيانـات المعجميـة وصيانتهـا، والتكامـل مـع أنظمـة معالجـة اللغـة الطبيعيـة. وتتوافـق الوثيقـة مـع معاييـر (XML) و(Unicode) كمـا تتكامـل مـع معيـار (ISO 24613-1:2024)، ويمكـن ربطهـا بعـدد مـن وثائـق المعاييـر الأخـرى، مثـل معاييـر TMF لإدارة المصطلحـات ومعاييـر الترجمـة الآليـة، وتهـدف هـذه المعاييـر إلى تحسـين جـودة وتناسـق المعاجـم وتسـهيل التبـادل الـدولي للمعلومـات المعجميـة، مـع تقليـل تكاليـف الإنتـاج والصيانـة، وتسـهيل التعـاون بـين الناشـرين، وتحسـين إمكانيـة الوصـول للمسـتخدمين.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

ومـن المحـاولات المبكـرة في وثائـق معايـير المعاجـم، دليـل ترمـيز البيانـات المعجميـة والدلاليـة الـذي طـوره مجموعـة دوليـة مـن المتخصصـين ( EAGLES Lexicon Interest Group)، ويتضمـن إطـاراً مرجعيًّـا شـاملاً يهـدف إلى توحيـد معايـير ترمـيز المعلومـات المعجميـة والدلاليـة في المعاجـم الإلكترونيـة وتطبيقـات معالجـة اللغـة الطبيعيـة. ويتمـيز الدليـل بمكوناتـه الأساسـية الـتي تشـمل تصنيـف العلاقـات الدلاليـة، وهيكلـة المداخـل المعجميـة، وترميز المعلومـات النحويـة والصرفيـة، وتمثيـل العلاقـات المعجميـة. كما يتسـم بالمرونـة في التطبيـق وقابليـة التوسـع والتوافـق مـع المعايـير الدوليـة؛ ممـا يسـهل تبـادل المـوارد المعجميـة بـين التطبيقـات المختلفـة ودعـم معالجـة اللغـات المتعـددة. وانتشـر اسـتخدام الدليـل -انتشـاراً واسـعاً- في تطويـر المعاجـم الإلكترونيـة وأنظمـة معالجـة اللغـة الطبيعيـة وأدوات الترجمـة الآليـة وتطبيقـات اسـتخراج المعلومـات.

ويمثـل الإطـار النظـري الشـامل لتطويـر المعاجـم متعـددة اللغـات (SIMPLE) -الـذي طـوره مجموعـة مـن المختصـين- نموذجـاً نظريًّـا شـاملاً لتطويـر المعاجـم متعـددة اللغـات، إذ يقـدم منهجيـة متكاملـة تجمـع بـين تقديـم نمـوذج موحـد لتمثيـل المعلومـات المعجميـة، وهيـكل معيـاري موحـد للمداخـل المعجميـة. ويتميـز الإطـار بتعـدد طبقاتـه الـتي تشـمل المعلومـات المعجميـة الأساسـية، والمعلومـات الدلاليـة، والعلاقـات بـين اللغـات، والمعلومـات النحويـة والصرفيـة. ويتسـم بالمرونـة العاليـة في تمثيـل الخصائـص اللغويـة وقابليـة التوسـع لدعـم لغـات جديـدة، مـع الحفـاظ علـى التوافـق مـع المعايـير الدوليـة. ويسـتخدم هـذا الإطـار اسـتخداماً واسـعاً في مجـالات متعـددة، تشـمل بنـاء المعاجـم متعـددة اللغـات، وتطويـر أنظمـة الترجمـة الآليـة، وتطبيقـات اسـترجاع المعلومـات، وأدوات معالجـة اللغـات الطبيعيـة، وغيرهـا.

ويتميـز النمـوذج المبكـر لتمثيـل المعاجـم الإلكترونيـة ( GENELEX)، بقابليتـه العاليـة لإعـادة الاسـتخدام في تطبيقـات متنوعـة. ويتكـون مـن عـدة مسـتويات هيكليـة تشـمل المسـتوى الصـرفي والتركيـبي والـدلالي، إضافـة إلى مسـتوى العلاقـات المعجميـة. ومـن

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً



أبـرز مميزاتـه: الفصـل بـين البيانـات وآليـات المعالجـة، وقابليـة التوسـع والتخصيـص، مـع الحفـاظ علـى التوافـق مـع المعايـير العالميـة ودعـم تعـدد اللغـات. واسـتُخدم النمـوذج استخداما واسـعاً في تطويـر المعاجـم الإلكترونيـة، وأنظمـة معالجـة اللغـات الطبيعيـة ونظـم الترجمـة الآليـة.

ويمثـل المسـتوى الـدلالي في نمـوذج (GENELEX) طبقـة معقـدة ومتكاملـة لمعالجـة المعـاني وعلاقاتهـا؛ إذ يتكـون مـن وحـدات دلاليـة أساسـية تشـمل المعـنى المركـزي والمعـاني الفرعيـة والعلاقـات بينهـا. ويعتمـد النمـوذج علـى شـبكة غنيـة مـن العلاقـات الدلاليـة، مثـل: الـترادف، والتضـاد، والاشـتمال، وعلاقـات الجـزء بالـكل، والتكافـؤ الـدلالي، ويتضمـن سـمات دلاليـة محـددة، تشـمل: الفئـات، والحقـول الدلاليـة، والقيـود الدلاليـة، والخصائص التمييزيـة. ويسـتخدم النمـوذج آليـات تمثيـل متقدمـة تتضمـن التمثيـل الهـرمي للمفاهيـم، والشـبكات الدلاليـة المترابطـة، والروابـط المرجعيـة بـين المعـاني وتمثيـل السـياقات الاسـتعمالية. وتُوظَّـف هـذه العناصـر في تطبيقـات معجميـة متنوعـة مثـل معالجـة تعـدد المعـاني وحـل الغمـوض الـدلالي واسـتنتاج العلاقـات الدلاليـة، كمـا تلعـب دوراً محـوريًّا في ربـط المعـاني بـين اللغـات المختلفـة؛ ممـا يجعـل هـذا المسـتوى ركـيزة أساسـية في بنـاء المعاجـم الآليـة القابلـة لإعـادة الاسـتعمال.

وفي المعاجـم المختصـة قدمـت المنظمـة الدوليـة للمعايـير آيـزو ISO تسـع وثائـق لمعايـير إدارة وضمـان جـودة المعالجـة الحاسـوبية للوحـدات المعجميـة المختصـة والمصطلحـات، وتتضمـن المعاجـم المختصـة طيفًـا واسـعًا مـن المعاجـم الـتي تمثـل اللغـة الاصطلاحيـة المسـتعملة لأغـراض خاصـة في المجـالات العلميـة والمهنيـة المتعـددة. وعنـد اسـتعراض معايير المعاجـم المختصـة أو المصطلحـات الـتي أصدرتهـا منظمـة الآيـزو، نجـد مجموعـة مـن الوثائـق الـتي اهتمـت بجوانـب مختلفـة مـن صناعـة وبنـاء معاجـم المصطلحـات وتخزينهـا إلكترونيـا وفقًـا لمعايـير مختلفـة، واشـتملت المعايـير على عـدد مـن الجوانـب المتصلـة بصناعـة وإعـداد المعاجـم المختصـة، ومنهـا:

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

١- معايير تعريف المصطلحات وربطها بطرق متعددة.

٢- تنظيم وتحليل الحقول الدلالية للمعاجم المختصة.

٣- معايير تصميم وبناء الأنظمة الحاسوبية الخاصة بمعالجة المداخل المعجمية المختصة.

٤- معايير الترميز اللغوي وغير اللغوي للمصطلحات.

٥- معايير تنظيم وإنتاج وصيانة قواعد بيانات المعاجم المختصة.

وتهدف أغلب المعايير والوثائق المرجعية التي صممت لتمثيل البيانات المعجمية إلى تحقيق مبادئ رئيسة أشار إليها (Wilkinson et al. 2016) ، وهي سهولة الوصول إليها، والبحث فيها، وكذلك معالجتها حاسوبيًّا معالجةً تكامليةً مع الموارد اللغوية المشابهة، وإعادة استعمالها لأغراض مختلفة، وإمكانية مشاركتها بسهولة في التطبيقات اللغوية الحاسوبية المختلفة.

وتعد معايير آيزو لإدارة قواعد بيانات المصطلحات والمعاجم المختصة (ISO 30042) من أهم الوثائق المتعلقة بمعايير تمثيل بيانات المعاجم الاصطلاحية الإلكترونية، وبُنيت على مجموعة كبيرة من الدراسات والبحوث المسحية الخاصة بمعرفة طرق إدارة ومعالجة بيانات المعاجم الاصطلاحية، ونشرت أول نسخة منها في عام (٢٠٠٩) وتوالت التحديثات المستمرة عليها حتى آخر إصدار لها عام (٢٠١٩). وعَرَّفت الوثيقةُ المعجمَ المختص بأنه: قاعدة بيانات تجمع فيها مجموعة من الكلمات والمصطلحات والمفاهيم التي تستعمل لأغراض خاصة، وتتضمن مداخلها المعجمية عددًا من التعريفات والمرادفات والعلاقات الدلالية المتنوعة. واعتمدت وثيقة المعايير على منهجية العلاقة الدلالية المنفردة لكل وحدة معجمية خاصة، وذلك من خلال ربط كل مدخل معجمي بمعنى ومفهوم اصطلاحي خاص، وفي حالة تعدد معاني المدخل فإنها تعتبر مداخل معجمية مستقلة، وهذا مخالف لظاهرة تعدد المعاني المعتادة في المداخل المعجمية

العامة. ومن المعايير التي اعتمدتها الوثيقة ويمكن تبنيها في الوثيقة الحالية، مجموعة من المعايير لضبط كتابة تعريف المصطلح الخاص.

وكان الدافع الرئيس لهذه الوثيقة هو تيسير عملية تبادل البيانات المعجمية الإلكترونية الخاصة بين اللغات والتخصصات المختلفة من خلال اعتماد مجموعة من المعايير العالمية لطريقة تنظيمها وتخزينها ومعالجتها في قواعد البيانات المحوسبة. ووفقًا لمعايير الوثيقة يتكون معجم المصطلحات من مجموعة من فئات البيانات التي تتضمن فئات إلزامية، مثل: المصطلح أو رأس المدخل المعجمي، واللغة، والتعريف، وفئات أخرى اختيارية، مثل: التحليل الصرفي، والحقل الدلالي، والتعليقات. وأوردت الوثيقة التمثيل المناسب لكل هذه البيانات في لغة (xml) التي تستعمل لتمثيل بيانات الموارد اللغوية.

## معايير بناء منصات تعليم اللغة العربية لغةً ثانيةً:

يُظهر تحليل الأدبيات ذات الصلة بوثائق معايير منصات تعليم اللغة عددًا من وثائق المعايير التي طورتها جهات مختلفة، تتضمن الفئات المتعددة التي يمكن أن تصنف إليها المعايير في هذا السياق، ويمكن الإفادة من كل هذه الأعمال في تطوير وثيقة شاملة لبناء منصات تعليم العربية لغةً ثانيةً.

ومن أهم وثائق المعايير التعليمية التي ينبغي الإفادة منها في الأطر والمعايير المرجعية لتعليم اللغات: معايير المجلس الأمريكي لتعليم اللغات الأجنبية، ومعايير الإطار المرجعي الأوروبي المشترك للغات؛ إذ تشكل هذه المعايير حجر الأساس في تطوير منصات تعليم اللغة الإلكترونية. فهي تقدم إطاراً شاملاً ودقيقاً لتصميم المسارات التعليمية، عبر تصنيف مستويات الكفاءة اللغوية إلى فئات متدرجة، مع تفصيل دقيق لخصائص كل مستوى. كما يمكن توظيف هذه المعايير في المنصات الإلكترونية عبر تصميم خوارزميات ذكية لتقييم مستوى المتعلم وتوجيهه إلى المسار المناسب (ACTFL, 2012; 2023)، (Council of Europe, 2023).

وتقدم معايير تيسول التقنية ( TESOL technology standards)، عددًا من المعايير التقنية التعليمية لتطوير منصات تعليم اللغة. وتتضمن إطارًا شاملًا لدمج التقنية في تعليم اللغة بطرق فعالة وشاملة. وتتضمن المعايير جوانب مهمة، مثل تصميم واجهات المستخدم، وتطوير الأدوات التفاعلية، وآليات التقييم الإلكتروني. ويمكن تطبيق هذه المعايير في تصميم أدوات التفاعل المباشر، مثل: المحادثات الصوتية، وغرف الدردشة التعليمية، والتدريبات التفاعلية. كما تقدم إرشادات حول كيفية توظيف التقنيات الحديثة، مثل: الذكاء الاصطناعي والواقع المعزز في تعليم اللغة. وتساعد هذه المعايير أيضًا في تطوير نظام متكامل للتغذية الراجعة الفورية وتتبع تقدم المتعلم (TESOL International Association, 2022; 2023).

وتشمل معايير (Web Content Accessibility Guidelines – WCAG) للوصول إلى محتوى الويب مجموعة من المبادئ الأساسية التي تضمن تيسير وصول جميع المتعلمين للمنصات التعليمية بسهولة. وتقدم هذه المعايير توجيهات شاملة لتصميم منصات تعليمية تلبي كافة احتياجات المستخدمين، بمن فيهم ذوو الاحتياجات الخاصة. وفي مجال تعليم اللغة، تتيح هذه المعايير خصائص متنوعة، مثل: خيارات تكبير النصوص، وضبط تباين الألوان، وإنشاء نسخ صوتية للمحتوى المكتوب والتوافق مع برامج قراءة الشاشة والتقنيات المساعدة الأخرى. ومن خلال تطبيق هذه المعايير، يصبح تعلم اللغة متاحاً للجميع، ومراعيًا للاحتياجات المختلفة للمتعلمين (World Wide Web Consortium, 2023; 2024).

كما تتضمن معايير الوصول العالمية أيضاً المبادئ الرئيسة لضمان إتاحة المنصة التعليمية لجميع المستخدمين؛ إذ تقدم هذه المعايير إرشادات تفصيلية حول كيفية تصميم منصات يمكن لجميع فئات المتعلمين الوصول إليها، بمن في ذلك ذوو الاحتياجات الخاصة. وفي سياق تعليم اللغة، تسهم هذه المعايير في تطوير مميزات عدة، مثل: تكبير النص، وتعديل التباين اللوني، وتوفير بدائل صوتية للمحتوى النصي. كما

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً



تضمن توافق المنصة مع قارئات الشاشة وغيرها من التقنيات المساعدة. وبذلك تجعل هذه المعاييرُ تعلمَ اللغة متاحًا للجميع، مع مراعاة احتياجات مختلف فئات المتعلمين (World Wide Web Consortium, 2023; 2024).

وهناك عدد من وثائق المعايير ذات الصلة بجودة التعليم الإلكتروني طورتها منظمة (Quality Matters)، وتهدف إلى تقديم إطار شامل لضمان جودة المحتوى التعليمي في المنصات الإلكترونية (Quality Matters, 2023). وتشمل ثمانية مجالات رئيسة للمعايير، مثل: معايير إعداد نبذة عن المقرر، وأهداف التعلم، والتقييم والقياس، والمواد التعليمية، وأنشطة المتعلم والتفاعل، وتقنيات المقرر، ودعم المتعلم، وإمكانية الوصول. وفي سياق منصات تعليم اللغة، تساعد هذه المعايير على تنظيم المحتوى تنظيماً منطقياً ومتسلسلاً، وتطوير أنشطة تفاعلية تدعم اكتساب المهارات اللغوية. كما تقدم توجيهات حول كيفية تصميم نظام دعم متكامل يساعد المتعلمين في تخطي الصعوبات وتحقيق أهدافهم التعليمية. كما قدمت ISTE (الجمعية الدولية للتقنيات التعليمية) عددًا من وثائق المعايير التي تركز في دمج التقنية في التعليم بفعالية. وتتضمن معايير شاملة للطلاب والمعلمين وقادة الجهات التعليمية والمدربين. وتهدف هذه المعايير إلى تعزيز الاستخدام المبتكر للتقنية في التعليم (International Society for Technology in Education, 2024).

وفيما يتعلق بالمعايير التقنية، نجد عددًا من الوثائق التي طورتها منظمات (IMS QTI, IEEE, SCORM)، وتركز في الأسس الفنية التقنية لتطوير منصات تعليم اللغة. حيث تركز معايير (IMS QTI) في جودة أدوات التقييم الإلكتروني وفعاليتها، في حين تركز معايير (IEEE) في البنية التحتية التقنية للمنصات التعليمية. وتُقدم المعايير التي طورتها (SCORM) إطاراً لتنظيم وتبادل المحتوى التعليمي بين المنصات المختلفة. كما تستعمل هذه المعايير في تطوير أنظمة تقييم آلي دقيقة للمهارات اللغوية، وواجهات

برمجية موحدة وقابلة للتطوير، وبناء نظام متكامل لتتبع تقدم المتعلم، وآليات لتبادل المحتوى التعليمي بين المنصات التعليمية (IMS Global Learning Consortium, 2024; IEEE LTSC, 2024).

وهناك عدد من وثائق المعايير التي طورتها (ISO) و(IEC) للتعلم الإلكتروني، وتقدم إطاراً شاملاً لضمان جودة وأمان المنصات التعليمية. وتشمل جوانب مهمة، مثل: أمن بيانات المتعلمين وخصوصيتها، ومعايير جودة المحتوى التعليمي، وآليات التفاعل والتواصل، وقابلية التوسع والتطوير للمنصات التعليمية. وفي سياق بناء منصات تعليم اللغة العربية لغةً ثانيةً، تساعد هذه المعايير على بناء منصات آمنة وموثوقة، مع ضمان حماية خصوصية المتعلمين وبياناتهم الشخصية (ISO/IEC 27001 ISO 40180).

وتقدم وثائق المركز الوطني للتعليم الإلكتروني إطاراً متكاملاً لتطوير التعليم الإلكتروني، وتشمل الوثائق معايير جودة التعليم الإلكتروني، وإطار المهارات الرقمية، ومعايير تطوير المحتوى الإلكتروني، ومواصفات منصات التعلم الإلكتروني. كما قدم مكتب التربية العربي مجموعة من وثائق المعايير لتعليم اللغة العربية وتطوير المحتوى الرقمي العربي. وتشمل الوثائق المعايير المهنية لمعلمي اللغة العربية، ومعايير تقويم البرامج التعليمية الإلكترونية، وإطار الكفايات التقنية للمعلمين.

وتشكل جميع وثائق المعايير -التي استعرضناها باختصار في هذا القسم- إطارًا متكاملًا لتطوير منصات تعليم اللغة، ويمكن الإفادة منها في تطوير وثيقة متكاملة وشاملة لبناء منصات تعليم العربية لغةً ثانيةً. ومن الضروري النظر إليها نظرة تكاملية؛ لأنها تتضمن جميع الجوانب التعليمية والتقنية والتنظيمية التي يلزم مراعاتها في تطوير منصات تعليم اللغة العربية لغةً ثانيةً؛ لضمان تطوير منصات تعليمية فعالة وقابلة للتطوير، تلبي احتياجات المتعلمين وتحقق أهداف تعليم اللغة.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً



## معايير بناء المدونات اللغوية وتوسيمها:

المدونة اللغوية (أو المدونة) هي مجموعة من المواد اللغوية الأصلية -غالباً على شكل نصوص- جُمِعت لدراستها آليا باستخدام الحاسب (Lakhouaja, 2018; Wynne & Prytz, 2005). وتبرز أهمية المدونات في كونها مصدرًا رئيسًا في الدراسات اللغوية، وفي مجال معالجة اللغة الطبيعية، وبناء نماذج الذكاء الاصطناعي. ويُراعى عند جمعها أن تبقى المواد اللغوية في شكلها الطبيعي دون إضافات جوهرية، باستثناء بعض البيانات الوصفية، مثل: وقت الجمع والمصدر.

وتضيف عملية توسيم المدونات معلومات لغوية مهمة غير موجودة في المدونة الخام، مثل: إضافة أقسام الكلام. ويمكن تقسيم العمل في بناء المدونات اللغوية **إلى نوعين:**

- بناء المدونة الخام: يتضمن جمع مواد لغوية بشكلها الطبيعي دون إضافات جوهرية لغوية إليها، باستثناء بعض البيانات الوصفية، مثل: وقت الجمع والمصدر.

- توسيم المدونات: يُضاف في المدوّنات الموسّمة معلومات لغوية جوهرية لا توجد في المدونة الخام، مثل إضافة أقسام الكلام إلى المدونة. ونظرًا إلى الاختلاف الرئيس في أصالة المادة اللغوية، فقد قُسّمت المعايير إلى فصلين؛ يتناول الفصل الخامس معايير بناء المدونات الخام، في حين يشمل الفصل السادس معايير التوسيم التي تتطلب إضافة تفاصيل لغوية إلى المدونة.

وقد مرّ مجال بناء المدونات اللغوية وتوسيمها بمراحل متعددة منذ نشأة اللسانيات الحاسوبية في منتصف القرن العشرين؛ فخلال الخمسينيات والستينيات، كان التركيز -بدايةً- في جمع نصوص لغوية خام دون إضافات لغوية جوهرية، مستفيدين من تطوّر الحاسب كأداة لتخزين البيانات ومعالجتها. ومع تقدّم الأبحاث اللغوية وخوارزميات

التحليـل الحاسـوبي، ظهـرت الحاجـة في التسـعينيات إلى توسـيم المدونـات (Pustejovsky & Stubbs, 2012; Ide, 1998) سـواء مـن حيـث أقسـام الـكلام أو التراكيـب النحويـة أو المعلومـات الصرفيـة، إلى أن صـار التوسـيم مطلبًـا رئيسًـا لتطبيقـات الـذكاء الاصطناعـي ومعالجـة اللغـة الطبيعيـة.

ومـع انتشـار الإنترنـت في بدايـة الألفيـة الثانيـة أصبحـت البيانـات متاحـة بسـهولة لاسـتخدامها في النمـاذج الإحصائيـة، ولا سـيما في الترجمـة الآليـة. ومـع تحـول معالجـة اللغـة مـن النمـاذج المعتمـدة علـى القواعد إلى نمـاذج التعلـم الآلي المعتمـدة على الـذكاء الاصطناعي، ازدادت الحاجـة إلى مدونـات موسـومة لتدريـب هـذه النمـاذج وتحسـين أدائهـا؛ ممـا أدى إلى كـثرة الدراسـات والجهـود العمليـة في بنـاء المدونـات وتطويـر وثائـق مرجعيـة لهـا.

بنـاء علـى ذلـك؛ ظهـرت الحاجـة إلى معايـير مرجعيـة لضبـط عمليـة بنـاء المدونـات وتوسـيمها لتحقـق أكـبر فائـدة. وتنقسـم هـذه المعايـير إلى نوعـين، بحسـب الغـرض الـذي صممـت بدايـة مـن أجلـه:

١. معايـير مختصـة بمدونـة معينـة: معايـير صُممـت لمدونـة معينـة، دون قصـد اسـتخدامها في مشـروعات أخـرى، مثـل المعايـير الخاصـة ببنـك بـين الشـجري Penn Treebank (Marcus et al., 1993).

٢. معايـير مصممـة للتوافقيـة: وهـي الـتي صممـت خصيصًـا لتكـون مرجعيـة للتوافقيـة بـين مشـروعـات بنـاء المدونـات وتوسـيمها، مثـل معيـار ترميـز المدونـات (Corpus Encoding Standard CES (Ide), 1998).

ومـن المهـم الإشـارة إلى أن النـوع الثـاني مـن المعايـير تأثـر تأثـراً كبـيراً بوجـود النـوع الأول؛ فقـد أسـهمت الوثائـق والمسـتندات المنشـورة للمدونـات الموسـومة -مثـل بنـك بـين الشـجري (Penn Treebank)- في تطويـر معايـير توسـيم المدونـات؛ لمـا تحتويـه هـذه الوثائـق عـادة مـن تفاصيـل حـول بنـاء المدونـات، وعمليـة التوسـيم، مثـل: اختيـار الوسـوم المناسـبة، وطريقـة

كتابة المستندات، وغيرها. وعلى ذلك، فإن المعايير من النوع الثاني تعتمد اعتماداً رئيساً على ما قدمته المعايير الخاصة بالمدونات المعينة.

وفيما يتعلّق باللغة العربية، فقد ظهرت جهودٌ مهمّة في بناء مدونات عربية موسومة، مثل البنك الشجري العربي (Arabic Treebank) الذي نفّذته جمعية البيانات اللغوية (LDC) بجامعة بنسلفانيا، ومدونة القرآن الكريم العربية التي تُقدّم تحليلات صرفية ونحوية ودلالية للقرآن الكريم، ومدونات موجّهة لتوسيم أسماء الأعلام، مثل مدونة (CAraNER) ومدونة (Wojood). غير أن أغلب هذه الأعمال يعاب عليها غياب دليل موحّد أو معيار تفصيلي يشرح الخطوات المرجعية لتوسيم المدونات العربية؛ مما دفع جهات عدة إلى نشر" أدلة توسيم" خاصة بها دون توحيد الجهود.

ويُعَد كتاب "المدونات اللغوية العربية: بناؤها وطرائق الإفادة منها" (العصيمي، ٢٠١٥) من أبرز المؤلّفات التي تناولت المدونات العربية، إذ عرض في خمسة مباحث قضايا تتعلّق بطرق بناء المدونات الخام والإفادة منها، ومدونات المعلّمين والمتعلّمين، وتطبيقات المدونات في الصحافة العربية، وصولًا إلى الرهانات المستقبلية في مجال بناء المدونات الحاسوبية وتوسيمها. ومع ذلك، لا تزال الحاجة ماسةً إلى أدلة معيارية متخصصة تُفرّق بوضوح بين معايير بناء المدونات الخام (مثل: اختيار النصوص وتصنيفها وتوثيقها) ومعايير التوسيم اللغوي (مثل: إضافة أقسام الكلام، والتحليلات النحوية والصرفية، وأسماء الأعلام)، بما يمكّن المطوّرين والباحثين من اتباع مرجعية موحّدة في بناء مشروعاتهم اللغوية. ومن هنا تبرز أهمية إعداد وثيقة جامعة للمعايير الخاصة ببناء وتوسيم المدونات باللغة العربية، ولا سيما في ظل قلة المصادر حولها.

كما استعرض (Zeroual ,2019) تاريخ المدونات العربية منذ الكتابة اليدوية "ما قبل الإلكترونية" وحتى النماذج الحالية في العصر الرقمي. وتنوّعت الأمثلة بين مدونات ضخمة، مثل: (ArTenTen) و(مدونة المائة مليار) تجمع نصوص الإنترنت، ومدونات أكثر تخصصًا كمدونة (أبي الخير للنصوص الإخبارية)، ومدونة (المقالات العربية) التي

تحـوي٥٠٠ ألـف مقالـة مصنّفـة، ومدونـة (فلـك) الخاصـة بمجمـع الملـك سـلمان العالـمي للغـة العربيـة. وعلـى الرغـم مـن وفـرة البيانـات، فـإن الباحثـين يلحظـون غيـاب الضوابـط المعياريـة لـدى كثيـرمـن هـذه المشـروعات؛ ممـا أسـهم في انتشـار بيانـات "مشوّشـة" لا تصلح للاسـتخدام المباشـر في تطبيقـات الـذكاء الاصطناعـي. لـذا بـات مـن الضـروري اعتمـاد معايـير محـددة في جمـع النصـوص الخـام وتنقيتهـا، وأخـرى موجّهـة إلى توسـيم هـذه النصـوص بمـا يضمـن الانسـجام في البنـاء اللغـوي ويهيّئ البيانـات للاسـتخدام الحاسـوبي.

تؤكـد الدراسـات السـابقة أنّ مرحلـة بنـاء المدونـات الخـام هـي أسـاس أي مشـروع لغـوي لاحـق، في حـين يُعَـدّ توسـيم المدونـات خطـوة جوهريـة في تطويـر نمـاذج الـذكاء الاصطناعـي ومعالجـة اللغـة الطبيعيـة. وبالرغـم مـن وجـود محـاولات لوضـع معايـير دوليـة –مثـل: (CES، TEI، LAF)، أو جهـود بحثيـة عالميـة (بنـك بـين الشـجري)– فـإنّ اللغـة العربيـة لا تـزال بحاجـة إلى أدلـة مفصّلـة تُراعـي خصوصياتهـا، وتقـدّم حلـولًا للتحديـات الصرفيـة والنحويـة والدلاليـة، سـواء في بنـاء المدونـات الخـام أو في توسـيمها. ومـن هـذا المنطلـق تنبثـق أهميـة الوثيقـة الراهنـة الـتي تسـتلهم المعايـير العالميـة، وتعيـد صياغتهـا بمـا يلائـم البيئـة العربيـة، وتقدمهـا في هيكليـة شـاملة تسـاعد الباحثـين والمطوّريـن علـى تأسـيس مدونـات متقنـة وموسـومة، ترتقـي بمعالجـة اللغـة العربيـة إلى مسـتويات عالميـة في الدقـة والاتسـاق.

## معايير دعم اللغة العربية في التطبيقات الحاسوبية:

تُظهـر مراجعـة تاريـخ دعـم اللغـة العربيـة في التطبيقـات الحاسـوبية أنّ البدايـة ارتبطـت بتطويـر أدوات الكتابـة علـى الآلـة الكاتبـة في القـرن التاسـع عشـر، ثـمّ تطـوّرت مـع ظهـور الحواسـيب وانتشـارها في منتصـف القـرن العشـرين، ومـا صاحـب ذلـك مـن تحديـات في ترميـز الحـروف العربيـة وتوزيعهـا علـى لوحـات المفاتيـح. وقـد دفعـت تلـك التحديـات إلى اقـتراح معايـير مبكّـرة، مثـل (ASMO 449) الخـاص بالمنظمـة العربيـة للتوحيـد القيـاسي (آزمـو / ASMO)، ثـم تسـجيله في المنظمـة الدوليـة للمعايـير (آيـزو) تحـت رمـز (ISO 9036).

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً



كما جرى تطوير معيار آخر تحت اسم (ASMO 708) لدعم الأحرف العربية واللاتينية معًا، سُجِّل لاحقًا بوصفه (ISO 8859-6).

ولقد مثّل ظهور معيار الترميز الموحّد (1991) (Unicode) نقطة تحوّل في دعم العربية واللغات الأخرى رقميًّا. فقد قدّم اتحاد يونيكود (Unicode Consortium) حلًّا جذريًّا لإشكالات تعدّد نظم الترميز وتعارضها، عبر تخصيص رمز فريد لكل حرف في جميع اللغات. وقد أتاح هذا المعيار إمكانية معالجة النص العربي بجميع تشكيلاته واتجاهه الكتابي من اليمين إلى اليسار. وعلى هذا الأساس، تبنت معظم أنظمة التشغيل الحديثة ولغات البرمجة الشائعة الترميز الموحّد؛ مما يسّر تطوير تطبيقات تدعم لغات متعددة -وفي مقدّمها اللغة العربية- دون الحاجة إلى إعادة هيكلة شاملة في بنية التطبيق.

إلى جانب جهود الترميز، ظهرت معايير ودراسات خاصة بواجهات المستخدم (UI) وتجربة المستخدم (UX) ركزت في ضرورة مراعاة الجوانب الثقافية واللغوية في تصميم التطبيقات. فمثلًا، خصّصت شركة ( Oracle) في دليلها "Oracle Application Server Reports Services" فصلًا كاملًا لدعم العولمة والدعم ثنائي الاتجاه (RTL)، لتوضيح كيفية بناء تطبيقات قادرة على الانتشار عالميًّا تراعي خصائص اللغة العربية من حيث اتجاه النص وأحرفه. وتوازت هذه الجهود مع وثائق أخرى، مثل "AIX Globalization" الصادرة عن (IBM)، التي تقدم إرشادات لتطوير تطبيقات قادرة على تكييف نفسها أثناء وقت التشغيل مع مختلف اللغات والثقافات، بما في ذلك اللغات المكتوبة من اليمين إلى اليسار -مثل العربية-.

وقد ظهرت معايير ومبادرات أوسع تُعنى بالواجهة الثقافية للتطبيقات. فعلى سبيل المثال، يُقدّم المعيار الدولي (ISO/IEC 15897:2011) إطارًا لتسجيل العناصر الثقافية الخاصة بواجهات المستخدم، بما فيها إعدادات اللغة والتقويم والعملات، وهو ما يحتاجه المطوّرون عند بناء تطبيقات تدعم الثقافات المختلفة. كما يتيح المستودع المشترك للبيانات المحلية (CLDR) الذي يُديره اتحاد يونيكود، مجموعةً غنيّة من

البيانـات (كالأرقـام والتواريـخ والعمـلات) لأكـثر مـن ٤٠٠ لغـة، يُحـدّث دوريًّـا لضمـان تلبيـة متغـيرات اللغـات والثقافـات. وتعتمـد شـركات عالميـة كـبرى، مثـل (Apple وGoogle وIBM وMeta وMicrosoft) علـى (CLDR) لدعـم اللغـات المختلفـة في أنظمتهـا ومنتجاتهـا المتعـدّدة.

وبالرغـم ممـا أُحـرز مـن تقـدّم ملحـوظ في دعـم العربيـة عـبر معايـير الترمـيز -مثـل: Unicode)، والمبـادرات الثقافيـة، مثـل: (CLDR) و ISO/IEC 15897))- فإنّـه مـا يـزال هنـاك فـراغ في وجـود وثيقـةٍ جامعـة تضـمّ هـذه الجوانـب كافـة؛ ابتـداءً مـن ترميـز الحـروف واتجـاه النصـوص، مـرورًا بخصائـص الكتابـة العربيـة وتشـكيلها، وانتهـاءً بالظـروف الثقافيـة (التقويـم الهجـري، العمـلات المحليـة، ...إلـخ).

لذلـك يظـلّ إصـدار إرشـادات حديثـة وشـاملة للمطوّريـن -تأخـذ في الحسـبان الخصوصيـة اللغويـة والثقافيـة العربيـة- هدفًـا محوريًّـا لتوفـير تجربـة رقميـة ثريـة تلـبي احتياجـات المسـتخدم العـربي علـى النحـو الأمثـل.

## المراجع :


- العصيمي، صالح فهد. (٢٠١٥). المدونات اللغوية العربية: بناؤها وطرائق الإفادة منها، مركز خدمة اللغة العربية، الرياض.
- الغامدي، سلطان عبدالله، وفلاته، أحمد بن محمد بن إبراهيم. (٢٠٢٣). معايير مقترحة لتصميم المنصات التعليمية الإلكترونية في ضوء المبادئ التوجيهية للنفاذ إلى محتوى الويب لذوي إعاقة اضطرابات النطق واللغة. دراسات عربية في التربية وعلم النفس، ع١٤٥، ١٥٩ – ١٨٢.
- القاضي، هشام بن صالح. (٢٠١٩). اكتساب العربية لغة ثانية بواسطة تطبيقات الجوال دراسة تقييمية لتخطيط أفضل، مجلة التخطيط والسياسة اللغوية، مركز الملك عبد الله بن عبد العزيز الدولي لخدمة العربية، الرياض، السنة الرابعة، عدد ٨.
- مكتب التربية العربي. (٢٠٢٣). المعايير المهنية لمعلمي اللغة العربية. مكتب التربية العربي لدول الخليج.
- المركز الوطني للتعليم الإلكتروني. (٢٠٢٠). معايير التعليم الإلكتروني للتدريب بالمملكة العربية، مجلس إدارة المركز الوطني للتعليم الإلكتروني، السعودية.

- ACTFL. (2012). *ACTFL proficiency guidelines.* American Council on the Teaching of Foreign Languages.
- ACTFL. (2023). *ACTFL performance descriptors for language learners*. American Council on the Teaching of Foreign Languages.
- Bunt, H., Alexandersson, J., Carletta, J., Choe, J.-W., Fang, A. C., Hasida, K., & Traum, D. (2010). Towards an ISO standard for dialogue act annotation. *In Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)* (pp. 2548–2555).

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً


- Council of Europe. (2023). *Common European Framework of Reference for Languages: Learning*, teaching, assessment (CEFR). Cambridge University Press.
- Farghaly, A., & Shaalan, K. (2009). Arabic natural language processing: Challenges and solutions. *ACM Transactions on Asian Language Information Processing (TALIP),* 8(4), 1–22.
- Habash, N. Y. (2010). *Introduction to Arabic natural language processing*. Morgan & Claypool Publishers.
- Ide, N. (1998). Corpus encoding standard: SGML guidelines for encoding linguistic corpora. *In Proceedings of LREC* (pp. 463–470).
- Ide, N., & Pustejovsky, J. (2017). Introduction: The handbook of linguistic annotation. In N. Ide & J. Pustejovsky (Eds.), *Handbook of linguistic annotation* (pp. 1–18). Springer. https://doi.org/10.10071_2-0881-024-94-978/
- IEEE LTSC. (2024). IEEE *standard for learning technology systems architecture*. Institute of Electrical and Electronics Engineers.
- IMS Global Learning Consortium. (2024). *IMS Question & Test Interoperability (QTI) specification* (Version 3.0). IMS Global.
- International Organization for Standardization. (2001). *ISO 15188: Project management guidelines for terminology standardization*. ISO.
- International Organization for Standardization. (2007). *ISO 860: Terminology work—Harmonization of concepts and terms.* ISO.
- International Organization for Standardization. (2008). *ISO 22128: Terminology products and services—Overview and guidance*. ISO.
- International Organization for Standardization. (2014). *ISO 241561-: Graphic notations for concept modelling in terminology work and its relationship with UML—Part 1: Guidelines for using UML notation in terminology work.* ISO.

- International Organization for Standardization. (2015a). *ISO 17100: Translation services—Requirements for translation services*. ISO.
- International Organization for Standardization. (2015b). *ISO 9000: Quality management systems—Fundamentals and vocabulary*. ISO.
- International Organization for Standardization. (2019a). *ISO 1087: Terminology work and terminology science—Vocabulary*. ISO.
- International Organization for Standardization. (2019b). *ISO 261621-: Management of terminology resources—Terminology databases—Part 1: Design*. ISO.
- International Organization for Standardization. (2019c). *ISO 30042: Management of terminology resources—TermBase eXchange (TBX)*. ISO.
- International Organization for Standardization. (2021). *ISO 126161-: Terminology work in support of multilingual communication—Part 1: Fundamentals of translation-oriented terminography*. ISO.
- International Organization for Standardization. (2022a). *ISO 126201-: Management of terminology resources—Data categories—Part 1: Specifications. ISO.*
- International Organization for Standardization. (2022b). *ISO 704: Terminology work—Principles and methods*. ISO.
- International Organization for Standardization. (2022c). *ISO SMART standardization proposal [Unpublished].* ISO/TC 37.
- International Organization for Standardization. (2023). *ISO/IEC 40180:2023 Information technology—Quality for learning, education and training—Fundamentals and reference framework.* ISO.
- International Society for Technology in Education. (2024). *ISTE standards for students*. ISTE.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً


- Leech, G. (1993). Corpus annotation schemes. *Literary and Linguistic Computing,* 8(4), 275–281. https://doi.org/10.1093/llc/8.4.275
- Maamouri, M., Bies, A., Buckwalter, T., & Mekki, W. (2004, September). The Penn Arabic Treebank: Building a large-scale annotated Arabic corpus. *In NEMLAR conference on Arabic language resources and tools* (Vol. 27, pp. 466–467).
- Marcus, M., Santorini, B., & Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics, 19(*2), 313–330.
- Nivre, J., Abrams, M., Agić, Ž., Ahrenberg, L., Antonsen, L., Aranzabe, M., ... Žabokrtsky, Z. (2020). Universal Dependencies v2: A multilingual treebank collection. In Proceedings of the 12th Language Resources and Evaluation Conference *(LREC 2020)* (pp. 4027–4036).
- Nivre, J., De Marneffe, M.-C., Ginter, F., Goldberg, Y., Hajič, J., Manning, C. D., ... Zeman, D. (2016). Universal Dependencies v1: *A multilingual treebank collection. In Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016)* (pp. 1659–1666).
- Palmer, M., & Xue, N. (2013). Linguistic annotation. In A. Clark, C. Fox, & S. Lappin (Eds.), *The handbook of computational linguistics and natural language processing* (pp. 238–258). Wiley-Blackwell.
- Pustejovsky, J., & Stubbs, A. (2012). *Natural language annotation for machine learning: A guide to corpus-building for applications.* O'Reilly Media.
- Quality Matters. (2023). *Higher education course design rubric* (7th ed.). Quality Matters.
- Šimunić, D., & Pavić, I. (2020). *Standards and innovations in information technology and communications*. Springer. https://doi.org/10.10074-44417-030-3-978/
- TESOL International Association. (2022). *Standards for technology in language teaching*. TESOL Press.

- TESOL International Association. (2023). *Technology standards framework for language teaching*. TESOL Press.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data, 3,* Article 160018. https://doi.org/10.1038/sdata.2016.18
- World Wide Web Consortium. (2023). *Web Content Accessibility Guidelines (WCAG)* 2.2. W3C.
- World Wide Web Consortium. (2024). *Making content usable for people with cognitive and learning disabilities* (W3C Working Draft). W3C.
- Wynne, M., & Prytz, Y. B. (2005). *Corpus approaches to the language of literature*. Retrieved April 12, 2012.
- Zeroual, I., & Lakhouaja, A. (2018). Data science in light of natural language processing: *An overview. Procedia Computer Science, 127*, 82–91.

# الفصل الثاني

## معايير بناء المنصات الرقمية لتعليم اللغة العربية لغة ثانية

- ملخص
- مقدمة
- مدخل تعريفي
- الدراسات السابقة
- نطاق المعايير وحدودها
- قائمة المعايير
- المراجع



معايير الحوسبة اللغوية العربية

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

## ملخص

يقـدم هـذا الفصـل مجموعـة مـن المعايـير الموضوعيـة لبنـاء منصـات رقميـة لتعليـم اللغـة العربيـة لغـة ثانيـة، مسـتهلًا بمقدمـة حـول أهميـة توظيـف التقنيـة والمعايـير في تعليـم اللغـة وتعلمهـا، مـع مدخـل تعريفـي موجـز بمفهـوم المنصـات الرقميـة ومكوناتهـا وأهـم أنواعهـا. كمـا يسـتعرض أهـم المبـادئ الـتي ينبغـي مراعاتهـا في برامـج التعلـم الرقـمي عمومـاً، ثـم يناقـش أهـم الدراسـات السـابقة الـتي ركـزت في إعـداد معايـير وأطـر عامـة لتطويـر المنصـات الرقميـة، سـواء في مجـال التعليـم عامـة أم في تعليـم اللغـة الثانيـة خاصـة. ويختتـم الفصـل بعـرض قائمـة تشـتمل علـى أهـم المعايـير الـتي يجـب مراعاتهـا في بنـاء منصـات تعليـم اللغـة العربيـة لغـة ثانيـة، وتطويرهـا.

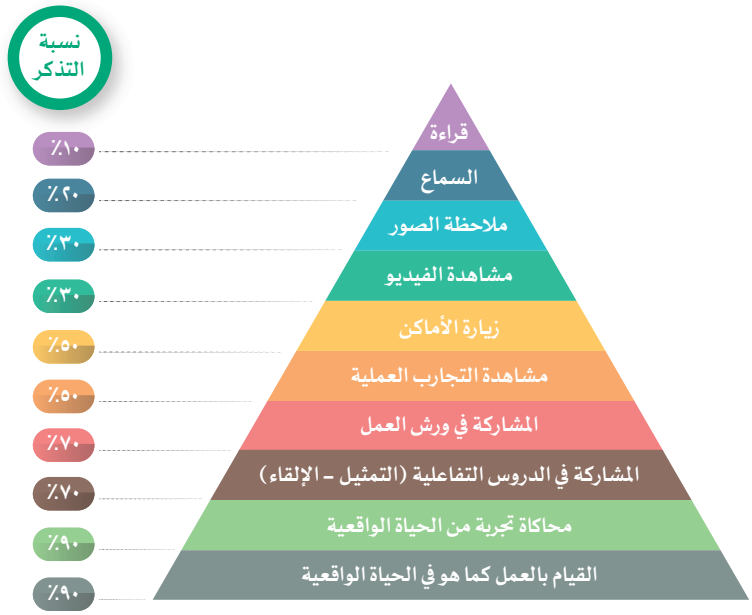
## مقدمة

أصبـح اسـتعمال التقنيـة عـن طريـق أدواتهـا ومنصاتهـا المختلفـة جـزءًا رئيسًـا في الحيـاة اليوميـة لكثـير مـن المعلمـين والمتعلمـين. وهـذا الاسـتعمال المكثـف للتقنيـة يؤكـد ضـرورة توظيفهـا في التعليـم عامـة وفي تعليـم اللغـة وتعلمهـا خاصـة. ويتفـق المعلمـون والباحثـون ومطـورو المناهـج التعليميـة علـى أن الدافعيـة عامـل رئيـس ومحـوري في تحديـد مسـتوى النجـاح في تعلـم اللغـات؛ ومـن ثـمَّ فـإن التقنيـات الحديثـة تمثـل وسـيلةً فعالـة لزيـادة دافعيـة المتعلمـين وتحفيزهـم علـى تطويـر مهاراتهـم اللغويـة. عـلاوة علـى ذلـك، فـإن توظيـف التقنيـة الحديثـة يُوجـد بيئـة تعليميـة غنيَّـة بالمصـادر التفاعليـة تدعـم اسـتقلالية المتعلـم وتعززهـا. (عثمانوفيـك، ٢٠٢٣).

وتشـير التقاريـر إلى أن اسـتخدام الحاسـوب في التعليـم كان أكـثر تأثـيرا في فئـة الصغـار والطـلاب ذوي القـدرات المحـدودة. ومـن هنـا تـبرز الحاجـة إلى دراسـة كيفيـة تسـخير التقنيـة لتحقيـق الأهـداف التعليميـة الخاصـة بتعلـم مهـارات اللغـة الثانيـة وعناصرهـا، وذلـك مـن خـلال تطبيـق التعليـم بمسـاعدة الحاسـوب، عـبر مشـروعات ومهمـات قائمـة علـى التقنيـة يقـوم بهـا متعلمـو اللغـة. كمـا أن

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

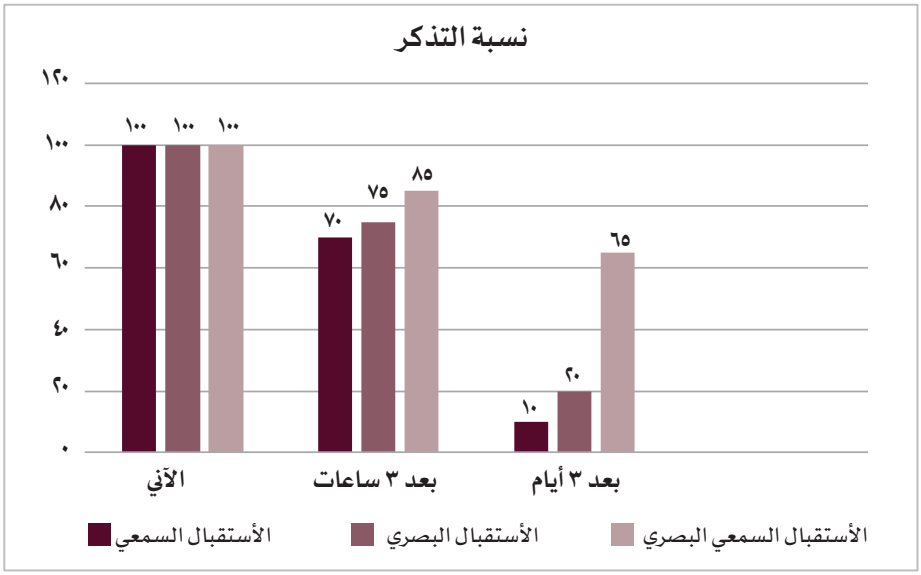
توظيف التقنية يعزز استمرارية أثر تعلم اللغة؛ إذ يساعد المتعلم على ممارسة المهارات اللغوية وتطبيقها في سياقات متعددة. ويوضح (الشكل١) أن ممارسة المهارة وتطبيقها إحدى أفضل الوسائل للتذكر واستدامة أثر التعلم.




الشكل ١: هرم نسبة تذكر المعلومات من خلال استعمال طرق تعلم متعددة.

وفي هذا السياق، يمكن القول: إنّ توظيف التقنيات المختلفة بسماتها المتنوعة يسهم إسهاماً فعالاً في تحسين طرق تعلم اللغة الثانية وزيادة فعاليتها، وذلك عن طريق "وسائط العرض التي تجعل من المعلومات الأكثر أهمية في الموقف التعليمي بارزة وواضحة؛ مما يجعل المتعلم فعالا بدرجة أكبر" (روبرت وآخرون، ٢٠٠٠: ٣٦٠). ويوضح (الشكل ٢) تأثير الوسائط المتعددة التي تقدمها التقنية في زيادة معدلات التذكر لدى المتعلمين، ويظهر تفوق الوسائط البصرية على الوسائط السمعية، كما يظهر أفضلية الجمع بينهما في تعزيز بقاء أثر التعلم.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً




الشكل ٢: رسم بياني يوضح أثر الوسائل السمعية والبصرية (حجازي، ٢٠٠٩).

وتـؤدي المعايـير والأطـر والوثائـق المرجعيـة دوراً رئيسـاً في تحقيـق أقـصى الفوائـد مـن اسـتخدام التقنيـات المختلفـة في تعلـم اللغـة الثانيـة؛ إذ إن هـذا الاسـتخدام لا يحقـق نتائجـه المرجـوة مـا لم يسـتند إلى مبـادئ مسـتقرة في تعلـم اللغـة الثانيـة وتعليمهـا، ويراعـي في الوقـت نفسـه أسـس بنـاء المنصـات الرقميـة التعليميـة وضوابطهـا. ومـن هـذا المنطلـق يسـهم هـذا الفصـل في تجسـير الفجـوة المعرفيـة في هـذا المجـال، عـبر تقديـم مجموعـة مـن المعايـير العامـة لبنـاء منصـات تعليـم اللغـة العربيـة لغـةً ثانيـةً.

ويؤكـد كل مـن "مينكـن، وهدسـون، وليـونج" (٢٠١٤) أن التعليـم المسـتند إلى المعايـير يؤثـر تأثـيرا إيجابيـا في النتـائج وممارسـات التقييـم (Al Fadda et al., 2020). وقـد أصبـح الهاتـف المحمـول -الـذي يملكـه ويسـتعمله أغلـب الطلبـة- يحتـوي على مجموعـة كبـيرة مـن الأدوات الـتي لـم يكـن يُتصـور اجتماعهـا معًـا في جهـاز واحـد (القـاضي، ٢٠١٩). ولهـذا يمكـن وصـف أجيـال المتعلمـين في الوقـت الحاضـر بـ"أجيـال التقنيـة" أو كمـا يصفهـم برينسـكي Prensky (٢٠٠١) بـ"جيـل التقنيـين

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

الأصيلين"، وهو الجيل الذي نشأ في بيئة غنية بالتقنية حتى غدت جزءًا أصيلًا من حياتهم، ومن ثم فإنّ استخدام التقنية في التعليم ضرورة تقتضيها ممارساتهم اليومية. وتؤكد هذه الحقيقة نتائج الدراسات البحثية التي تظهر أن متوسط الوقت الذي يقضيه الطلاب في استخدام التقنية في الولايات المتحدة قد تضاعف من قرابة ربع ساعة (٠,٣) يومياً في عام ٢٠٠٨ إلى أكثر من ثلاث ساعات (٣,٣) يوميا في ٢٠١٧م (القاضي، ٢٠١٩). كما شهد استعمال التطبيقات التعليمية في مطلع ٢٠٢٤، زيادة بنحو ٢٠ ٪ مقارنة بالمتوسط اليومي المسجل في ديسمبر ٢٠٢٣، في حين زاد خلال عام ٢٠٢٤ نمو استعمال تطبيقات تعلم اللغات خاصة بنسبة ٢٦٪، كما تظهر الأرقام الخاصة بإحصائيات عمليات التثبيت (Ceci, 2024).

وعلى الرغم من أن معظم تطبيقات تعلم اللغة العربية عبر الإنترنت أو الهواتف الذكية تعاني من ضعف المحتوى، وتفتقر إلى سمات التفاعلية والجاذبية وجودة التصميم، فإنها تحظى بشعبية كبيرة، وبعضها حقق ملايين التنزيلات أو المتابعات. ويدل هذا الإقبال على رغبة عدد كبير من المستخدمين في مختلف أنحاء العالم إلى تعلم اللغة العربية عبر الإنترنت والتقنيات الحديثة؛ ولذلك فالحاجة ملحة لإعداد معايير عالمية ووثائق مرجعية تسهم في تحسين جودة منصات وتطبيقات تعليم اللغة العربية وفاعليتها، وتكون مستندة إلى النظريات والمبادئ المستقرة في تعليم اللغة الثانية وتعلمها. لذلك تسعى هذه الوثيقة إلى الإسهام في تحقيق هذا الهدف عن طريق تقديم مجموعة من المعايير المهمة لبناء منصات تعليم اللغة العربية لغةً ثانيةً وتطويرها.

## مدخل تعريفي

يمكن تعريف المنصة الرقمية بأنها: "نظام تعليمي إلكتروني يُقدم عبر الإنترنت أو تطبيقات الحاسوب والهواتف الذكية، ويعتمد مبدأ التعلم المستقل أو المدمج الذي يجمع بين التعلم التقليدي في الصف مع المعلم والتعلم عبر الإنترنت". فيمكن للطالب استعمال المنصة بوصفها مصدرًا للتعلم الذاتي، كما يمكن للمعلم أو المدرب توظيفها لتعزيز عملية التعليم التقليدي في الصف باستخدام تقنيات التعليم المتاحة فيها.

أمـا التعليـم الإلكـتروني -بوجـه عـام- فيقصـد بـه: عمليـة تعليـم المعـارف والمهـارات والقيـم، وتعلمهـا باسـتخدام التقنيـات الحديثـة في قوالـب متنوعـة، مثـل: الحاسـوب والهاتـف المحمـول، وأجهـزة المسـاعد الرقـمي الشـخصي، والأجهـزة التقنيـة القابلـة للارتـداء، وغيرهـا مـن الأجهـزة والتقنيـات المسـتجدة لغرض التعليـم والتدريـب وإدارة المعرفـة. ويُعَرَّف التعلـم الإلكـتروني أيضـاً أنـه: "اسـتخدام تكنولوجيـا المعلومـات والاتصـالات (ICT) لتعزيـز ودعـم التعلـم في التعليـم العـالي" (OECD, 2005). ويتطلـب التعلـم الإلكـتروني "اسـتخدام الإنترنـت وتقنيـات أخـرى مهمـة لإنتـاج المـواد التعليميـة، وإدارة تعليـم المتعلمـين، وتنظيـم الـدورات في المؤسسـة" (Arkorful et al., 2014). ويمكـن تعريفـه كذلـك بأنـه: "طريقـة مبتكـرة لتوفـير بيئـة تعلـم جيـدة التصميـم وتفاعليـة وموجهـة للمتعلـم متاحـة لأي شـخص في أي مـكان وفي أي وقـت، باسـتخدام خصائـص ومـوارد الإنترنـت والتكنولوجيـا الرقميـة وفقًـا لمبـادئ التصميـم التعليـمي المناسـبة لبيئـة تعلـم مفتوحـة ومرنـة وموزعـة" (OECD, 2005). وقـد يُعـرف أيضـا بأنـه طريقـة للتعليـم تسـتخدم فيهـا التقنيـة بجميـع وسـائطها المتعـددة مـن صـوت وصـورة ورسـومات، وآليـات بحـث ومكتبـات إلكترونيـة، لإيصـال المعلومـة للمتعلـم بأقصـر وقـت وأقـل جهـد وأكـبر فائـدة (اليـوبي، ٢٠١٩)، وبهـذه الطريقـة يتفاعـل كلا الطرفـين في العمليـة التعليميـة عـن طريـق هـذه الوسـائط لتحقيـق أهـداف تعلـم محـددة (OECD, 2005). ويمكـن تعريـف التعليـم عـن طريـق التقنيـة بأنـه: مجموعـة متكاملـة مـن الخدمـات التفاعليـة عـبر الإنترنـت الـتي تقـدّم للمتعلمـين والمعلمـين وأوليـاء الأمـور وغيرهـم مـن المعنيـين بالتعليـم المعلومـات والأدوات والمـوارد الـتي تدعـم وتعـزز الخدمـات التعليميـة وإدارتهـا، وهـي نظـام شـامل يتيـح التدريـب الآمـن والتعلـم عـبر الإنترنـت والتعليـم الإلكتروني باسـتخدام واجهـة مسـتخدم مخصصـة (دحمـاني، ٢٠١٩).

وتتعـدد مكونـات المنصـة التعليميـة الإلكترونيـة، وتختلـف بحسـب الأجهـزة أو الطريقـة المسـتعملة في تقديمهـا؛ فتشـمل -علـى سـبيل المثـال-: أنظمـة إدارة تعلـم، والمـوارد التعليميـة المختلفـة، ومنصـات المجموعـات أو الفصـول الدراسـية. ويمكـن تصنيـف منصـات تعليـم اللغـة مـن زوايـا متعـددة، فيمكـن تصنيفهـا مـن حيـث طـرق الوصـول إليهـا إلى منصـات

تجاريـة تتطلـب رسـومًا أو ترخيصًـا لاسـتعمالها، ومنصـات مجانيـة يمكـن الوصـول إليهـا بـدون مقابـل. ويدخـل فيهـا مـا يعـرف بمنصـات التعلـم المفتـوح الـتي تشـارك في إعدادهـا مجموعـة مـن المؤسسـات التعليميـة، بهـدف تقديـم مجموعـة مـن البرامـج التعليميـة المهاريـة والأكاديميـة، وتسـهم مثـل هـذه المنصـات في تحقيـق المسـاواة بـين مختلـف فئـات المتعلمـين في الوصـول إلى الفـرص التعليميـة بـدون مقابـل.

وهنـاك **مجموعـة مـن المبـادئ العامـة الـتي يفضـل مراعاتهـا في تصميـم منصـات التعلـم الإلكتروني**، ويمكـن تصنيفهـا لمجـالات متعـددة، مـن أهمهـا:

- الموضـوع: موضـوع تعليـمي صالـح للتقديـم في صـورة واقـع افـتراضي، ويتطلـب اختيـار عنـوان يناسـب موضـوع التعلـم ويعكـس مضمونـه.
- الأهـداف: توضيـح الهـدف التعليـمي العـام ثـم الأهـداف الإجرائيـة وارتباطهـا بالموضـوع والمحتـوى التعليـمي.
- المحتوى: الدروس ومضامينها التعليمية ومدى ارتباطها بالموضوع وتحقيقها للأهداف.
- التغذية الراجعة: تقديم الدعم الفوري المناسب لتعزيز ردود الأفعال.
- الاسـتخدام: التميـز بسـهولة الاسـتخدام مـع الفعاليـة والكفـاءة في الاسـتجابة السـريعة لـردود أفعـال المسـتخدمين.
- الواجهـة: البسـاطة والخلـو مـن التعقيـد، واختيـار مـا يناسـب المتعلـم مـن حيـث الألـوان والأصـوات والأشـكال، وتيسـير الوصـول إلى أدوات تدعـم عمليـات البحـث والتفاعـل وتلقـي ردود الأفعـال.
- التفاعـل: التحفـيز والمشـاركة الفعالـة في إحـداث التعلـم، والاسـتمرارية في الأداء وتطويـر المهـارات التعلّميـة.
- التحكـم: التحكـم في الزمـن والوسـائط ومكونـات البيئـة عـن طريـق علامـات إرشـادية وتوضيحيـة للبحـث والتفاعـل.

- توفير عنصر مرونة الانتقال بين مكونات البرنامج وعناصر النظام من غير ضياع في البيئة التعليمية.
- المهمات: تحديد مهمات تعليمية واضحة تحقق الأهداف وترتبط بالموضوع، مع ضبط وقت الإنجاز والاستخدام.
- الأدوات: تتميز الأدوات بالمرونة، والتعدد، والجاذبية، بالإضافة إلى السلامة والاشتغال.
- الإدارة: نظام إدارة المحتوى، وتدبير التعلم، وطرق إدارة الاتصال والتقييمات والاختبارات.

وعند استعراض المواقع الإلكترونية لتعليم العربية للناطقين بغيرها تظهر لنا الحاجة إلى تعزيز مبدأ التفاعلية في هذه المواقع، عن طريق:

- الاعتناء بجودة التصميم.
- الاهتمام بعنصر الجاذبية.
- التناسق في عرض المحتوى.
- التنويع في الأنشطة والتدريبات.
- شمولية التطبيقات للمهارات.
- المرونة في التصفح.
- التجاوب التفاعلي مع المتعلمين والتغذية الراجعة الفورية (Heimo,2015).

وبحسب دراسة (Roschelle, 2003) ينبغي أن تسمح المنصة الإلكترونية للمعلمين بالحصول على ردود فعل من الطلاب في الصف؛ حتى يكونوا على علم بأداء الطلبة الفردي؛ فيجمع النظام إجابات الطلاب، وينشئ الجداول، أو الأنماط، أو الرسوم البيانية الأخرى إذا كان ذلك ممكنًا (الدراسات الإحصائية، أو الاستطلاعات، ...إلخ). وقد

حققت أنظمة الاستجابة في الفصل نتائج جيدة في التطبيق العملي؛ مما أدى إلى تحسين مشاركة الطلاب في الصف.

وذكر (Curtis et al., 2002) مجموعة من الخصائص الرئيسة التي ينبغي أن تتسم بها المنصة التعليمية، ومنها:

**الحفظ والتسجيل والاستمرارية:** تحفظ المنصة أعمال الدارسين في حساباتهم الشخصية عليها، وتمكنهم من المتابعة المستمرة لتطورهم.

**سهولة الوصول:** يمكن للمتعلمين الوصول إلى مستنداتهم وبياناتهم أو مقاطع الفيديو الخاصة بهم من أي مكان، فتتيح تلك المعلومات بناءً على طلباتهم؛ مما يجعل التعلم ذاتي الإدارة.

**الفورية:** يمكن للمتعلمين الحصول على أي معلومات فورًا أينما كانوا؛ مما يمكنهم من حل المشكلات بسرعة. وفي المقابل، يمكن للمتعلمين تسجيل الأسئلة والبحث عن الإجابات لاحقًا.

**التفاعلية:** يمكن للمتعلمين التفاعل المتزامن -وغير المتزامن- مع الخبراء والمعلمين أو الزملاء؛ مما يجعل الخبراء أكثر وصولًا، والمعرفة أكثر توافرًا.

**الأنشطة التعليمية:** ينبغي أن تتضمن الأنشطة التعليمية عرض المشكلات والمعرفة المطلوبة بأشكالها الطبيعية والأصلية. ويساعد ذلك المتعلمين على ملاحظة مميزات المواقف التي تجعل الإجراءات ذات الصلة. وحسب (تشين وآخرون، ٢٠٠٢) فإن أنظمة الجمع التعاوني للبيانات تمكّن المتعلمين من اكتساب خبرة واقعية وفهم أعمق لما يتعلمونه؛ إذ يتعامل المتعلمون مع الأشياء الحقيقية، ويلتقطون صورًا لها، ثم يعرضونها لاحقاً في الفصل. فعلى سبيل المثال، يتيح نظامٌ لمراقبة الطيور للطلاب التقاط صور للطيور خارج الفصل باستخدام أجهزة اليد، ومشاركة هذه البيانات مع المعلمين والطلاب الآخرين.

**التعلم التعاوني:** بغض النظر عن السيناريو التعليمي، يمكن لأجهزة التعلم المتنقلة مساعدة الأفراد في التعلم معًا؛ مما يقلل القيود ويعزز عملية التعلم.

**المرونة:** يتيح التعلم المتنقل للمستخدمين الفرصة ليكونوا مبدعين ومرنين؛ إذ يمكن إنشاء أنظمة تعلم متنوعة دون قيود للإبداع.

**البساطة والإمتاع:** يتيح التعلم المتنقل تجربة تعلم ممتعة تُكسب المتعلم شعورًا بالإنجاز؛ لكونه يتيح بيئة تعلم تفاعلية مرنة وقابلة للحركة. وتعزز البساطةَ فعاليةُ التعلم، إذ تقلل من التعقيدات وتزيد من الوقت المخصص للتعلم.

**قابلية الاسترجاع:** ينبغي أن تتيح المنصة التعليمية إمكانية استرجاع المادة التعليمية للطالب عند الحاجة، حتى يتمكن الطالب من مراجعة المادة متى شاء؛ الأمر الذي يجعل المنصة أكثر مرونة وقابلية للاستخدام.

وقد ذكرت وثيقة "معايير التعليم الإلكتروني للتدريب بالمملكة العربية السعودية" بعضَ المعايير المهمة للتعليم الإلكتروني، التي يمكن تبنيها والإفادة منها في هذه الوثيقة، ومن أهمها:

- توفير البنية التقنية اللازمة لتنفيذ برامج التدريب الإلكتروني.
- ضمان خصوصية بيانات المستفيد عن طريق وجود سياسات وإجراءات مصاغة بوضوح.
- توفير نظام دخول موحد وآمن.
- توفير أنظمة للتحقق من هوية المستفيد.
- توفير أنظمة إدارة التعلّم، والفصول الافتراضية والاختبارات الإلكترونية.
- أن تدعم الأنظمة مختلف أنواع الأجهزة باختلاف أنظمة تشغيلها وأحجام شاشاتها، بما في ذلك الهواتف الذكية والأجهزة اللوحية، كما تدعم جميع المتصفحات وقواعد البيانات.

- توفير تطبيقات على الهواتف الذكية لأنظمة التدريب الإلكتروني.
- وجود نظام تنبيهات إلكتروني لتنبيه المتدرب باستمرار على المهمات المطلوبة.
- أن تتيح الأنظمة خاصية البحث عن مختلف أنواع المحتويات الرقمية.
- تقسيم المحتوى إلى أجزاء صغيرة قابلة لإعادة الاستخدام.
- توفير المحتوى الرقمي بأشكال مختلفة (نصوص ومواد سمعية، ومواد مرئية) تدعم احتياجات المتدرب وخياراته المتنوعة.
- توفير تقييم ذاتي يمكن المتدرب من التحقق من تقدمه في التعليم، ويُقدّم له تغذية راجعة.
- يحدد الحد الأدنى من المعارف والمهارات التقنية والكفايات المطلوبة من المتدرب والمدرب للالتحاق بالتعليم الإلكتروني.

ومن أهم المهمات الخاصة بالمدرس داخل المنصة التعليمية، قيامه مع طلبته بالأدوار الآتية (الباوي وغازي، ٢٠١٩):

- تحديد أهداف الدروس والمقررات التي يسعى إلى تحقيقها.
- اختيار أو إعداد أساليب التقييم لقياس مدى تحقق الأهداف.
- متابعة حضور الطلبة ورصد تقدمهم الدراسي.
- تنظيم بيئة التعلم وجعلها مريحة وتعاونية.
- حث الطلبة على التعلم وتشجيعهم على المشاركة الفاعلة في الأنشطة الصفية.
- تكليف الطلبة بالقيام بالتدريبات والأنشطة والمشروعات.
- تنظيم التفاعل والنقاشات الصفية بينه وبين طلبته، وبين الطلبة بعضهم مع بعض.
- تقديم العون والإرشاد الأكاديمي للطلبة، وحل مشاكلهم الدراسية.
- إرشاد الطلبة إلى مصادر التعلم الإضافية المتاحة على الشبكة.

## الدراسات السابقة

تتضمن معايير بناء منصات تعليم اللغة إلكترونيًّا مجموعة متكاملة من المواصفات والمعايير التي يمكن تصنيفها إلى فئات متعددة، ومن أهمها:

- **المعايير التعليمية:** تشكل حجر الأساس في تطوير تطبيقات تعليم اللغة إلكترونيًّا؛ إذ تركز في صياغة أهداف تعليمية واضحة وقابلة للقياس تلبي احتياجات المتعلمين. وتضمن هذه المعايير التدرج المنطقي في عرض المحتوى من السهل إلى الصعب، مع توفير الدعم التعليمي المناسب في كل مرحلة. كما تؤكد أهمية التكامل بين المهارات اللغوية الأربع (الاستماع، والتحدث، والقراءة، والكتابة)، ودمج العناصر الثقافية التي تعزز فهم اللغة في سياقها الاجتماعي والثقافي. وتتضمن أيضاً آليات متنوعة للتقييم والتغذية الراجعة تساعد المتعلم على تتبع تقدمه وتحسين أدائه.
- **المعايير التقنية:** تتناول الجوانب الفنية والتشغيلية للمنصة؛ حتى تضمن سهولة الوصول إلى واجهة المستخدم وتوافقها مع مختلف الأجهزة والمنصات. وتركز في تحسين أداء النظام وضمان سرعة استجابته، مع الحفاظ على أمن وخصوصية بيانات المستخدمين. كما تشمل معايير موثوقية النظام وقدرته على العمل بكفاءة دون انقطاع، مع توفير تحديثات دورية تضمن مواكبة التطورات التقنية وإصلاح أي مشكلات قد تظهر أثناء الاستخدام.
- **معايير صناعة المحتوى اللغوي التعليمي:** تركز في جودة المادة اللغوية المقدمة وصحتها، مع مراعاة الدقة العلمية واللغوية. وتضمن هذه المعايير ملاءمة المحتوى ثقافياً للفئة المستهدفة، مع الحرص على ارتباطه بأهداف التعلم وحاجات المتعلمين. كما تشمل دمج الوسائط المتعددة دمجا فعّالاً يعزز التعلم، وتوفير عناصر تفاعلية تشجع على المشاركة النشطة. وتدعو إلى تنظيم المحتوى بطريقة منطقية ومتسلسلة تسهل عملية التعلم وتحقق الأهداف المرجوة.

- **معايـير تجربـة المسـتخدم:** تقـدِّم تجربـة تعلـم سلسـة وممتعـة، عـن طريـق تصميـم واجهـة سـهلة التنقـل وبديهيـة الاسـتخدام. وتتيـح خيـارات تخصيـص متنوعـة تمكـن المتعلـم مـن تكييـف التطبيـق وفـق تفضيلاتـه واحتياجاتـه. كمـا تتيـح أدوات لتتبـع التقـدم ومعالجـة الأخطـاء بفعاليـة، مـع تقديـم نظـام دعـم ومسـاعدة شـامل يسـتجيب لاستفسـارات المسـتخدمين ويسـاعدهم في حـل المشـكلات الـتي قـد تواجههـم.
- **معايـير إمكانيـة الوصـول:** تضمـن إمكانيـة إفـادة جميـع المسـتخدمين مـن المنصـة، بمـن فيهـم ذوو الاحتياجـات الخاصـة. وتشـمل قابليـة إدراك المعلومـات عـن طريـق توفـير بدائـل نصيـة للمحتـوى غـير النـصي، وقابليـة تشـغيل الواجهـة باسـتخدام مختلـف وسـائل الإدخـال، وقابليـة فهـم المحتـوى والتعليمـات بوضـوح. كمـا تضمـن جـودة التنفيـذ التقـني ودعـم التقنيـات المسـاعدة، مثـل: قارئـات الشاشـة، وأدوات التكبـير.
- **معايـير تقنيـات التعليـم:** تقـدم إطـارًا شـاملًا لدمـج التقنيـة في التعليـم؛ إذ تركـز في تعزيـز مشـاركة المتعلـم وتطويـر مهـارات المواطنـة الرقميـة. كمـا تدعـم بنـاء المعرفـة عـن طريـق التصميـم المبتكـر والتفكـير الحاسـوبي، وتشـجع التواصـل الإبداعـي والتعـاون العالـمي. وتهـدف هـذه المعايـير إلى إعـداد متعلمـين قادريـن علـى اسـتخدام التقنيـة بفعاليـة في تعلـم اللغـة والتواصـل مـع الآخريـن.
- **معايـير ضمـان الجـودة:** تضمـن اسـتمرارية تحسـين وتطويـر التطبيـق عـن طريـق المراجعـة الدوريـة للمحتـوى ومراقبـة الأداء. وتشـمل آليـات لجمـع وتحليـل تغذيـة المسـتخدمين الراجعـة واسـتخدامها في التحسـين المسـتمر. كمـا تؤكـد أهميـة التوثيـق الدقيـق لجميـع العمليـات وتوفـير خدمـات دعـم فعالـة تسـتجيب لاحتياجـات المسـتخدمين وتضمـن تجربـة تعلـم ناجحـة.
- **معايـير التقييـم:** تتنـاول جميـع جوانـب قيـاس تعلـم اللغـة؛ فتضمـن موثوقيـة أدوات التقييـم المسـتخدمة وصحتهـا. وتراعـي التـدرج في مسـتويات الصعوبـة بمـا

يتناسب مع مستوى المتعلمين وأهداف التعلم. كما تتيح أدوات لمراقبة تقدم المتعلم وتقدير إنجازاته، مع تقديم تغذية راجعة بناءة تساعده في تحسين أدائه وتحقيق أهدافه التعليمية.

وتُظهر مراجعة الوثائق المتعلقة بمعايير منصات تعليم اللغة جهوداً متعددة قامت بها مؤسسات وهيئات دولية متخصصة في التقنيات التعليمية، وتكشف هذه الوثائق عن تنوعٍ في الأطر المعيارية المعتمدة عليها، يمكن الإفادة منها في تطوير منصات تعليم العربية لغةً ثانيةً.

وتعد معايير المجلس الأمريكي لتعليم اللغات الأجنبية (ACTFL 2012, 2023). إطارًا مرجعيًا رئيسًا في مجال تقييم الكفاءة اللغوية وتطوير المناهج. وتتميز هذه المعايير بتقديمها تصنيفاً دقيقاً لمستويات الكفاءة اللغوية، من المبتدئ إلى المتميز، مع وصف تفصيلي لخصائص كل مستوى ومؤشرات الأداء المتوقعة. وتشكل هذه المعايير أساساً متيناً لتصميم المسارات التعليمية في المنصات الإلكترونية؛ إذ يمكن توظيفها في تطوير خوارزميات ذكية لتقييم مستوى المتعلم وتوجيهه للمسار المناسب. ويقدم الإطار المرجعي الأوروبي المشترك للغات (CEFR) نظاماً متكاملاً لتعليم وتقييم اللغات، ويصنف الكفاءة اللغوية إلى ستة مستويات رئيسة (A1-C2). ويتميز الإطار بتركيزه في المهارات التواصلية وقدرات المتعلم في استخدام اللغة في سياقات حقيقية. كما يقدم مؤشرات أداء تفصيلية لكل مستوى؛ مما يجعله أداة قيمة في تصميم وتقييم برامج تعليم اللغة عبر المنصات الإلكترونية (Council of Europe, 2023).

وتقدم معايير تيسول التقنية (TESOL Technology Standards) إطاراً شاملاً لدمج التقنية في تعليم اللغة. وتشمل هذه المعايير تصميم واجهات المستخدم، وتطوير الأدوات التفاعلية، وآليات التقييم الإلكتروني. كما تقدم إرشادات حول توظيف التقنيات الحديثة مثل الذكاء الاصطناعي والواقع المعزز في تعليم اللغة، وتطوير نظم التغذية الراجعة الفورية، وتتبع تقدم المتعلم (TESOL International Association, 2022; 2023).

تضمـن معايـيرُ الوصـول العالميـة (WCAG) وصـولَ جميـع المسـتخدمين إلى المنصـات التعليميـة، وتقـدم هـذه المعايـير توجيهـات مفصلـة لتصميـم منصـات متاحـة لجميـع فئـات المتعلمـين، ولا سـيما ذوي الاحتياجـات الخاصـة. وفي مجـال تعليـم اللغـة، تُسـهم هـذه المعايير في تطويـر خصائـص مهمـة، مثـل إمكانيـة تكبـير النصـوص، وضبـط تبايـن الألـوان، وتوفـير محتـوى صـوتي بديـل للنصـوص المكتوبـة (World Wide Web Consortium, 2023; 2024).

وطـورت منظمـة (Quality Matters) إطـاراً شـاملاً لضمـان جـودة التعليـم الإلكـتروني. ويحتـوي هـذا الإطـار علـى ثمانيـة مجـالات رئيسـة للمعايـير، تشـمل: نبـذة المقـرر، وأهـداف التعلـم، والتقييـم والقيـاس، والمـواد التعليميـة، وأنشـطة المتعلـم والتفاعـل، وتقنيـات المقـرر، ودعـم المتعلـم، وإمكانيـة الوصـول. وفي سـياق منصـات تعليـم اللغـة، تسـاعد هـذه المعايـير في تنظيـم المحتـوى وتطويـر الأنشـطة التفاعليـة وتصميـم نظـم الدعـم المتكاملـة (Quality Matters, 2023). وقدمـت الجمعيـة الدوليـة للتقنيـات التعليميـة (ISTE) معايـير شـاملة لدمـج التقنيـة في التعليـم، تضمنـت إرشـادات للطـلاب والمعلمـين وقـادة المؤسسـات التعليميـة والمدربـين، وتهـدف إلى تعزيـز الاسـتخدام المبتكـر للتقنيـة في العمليـة التعليميـة، وهـي تقـدم إطـاراً مرجعيـاً مهمـاً لتطويـر الكفايـات الرقميـة في منصـات تعليـم اللغـة (International Society for Technology in Education, 2024).

وفيمـا يتعلـق بالمعايـير التقنيـة، فقـد طـورت منظمتـا (IMS Global) و(IEEE) معايـير متخصصـة تركـز في الجوانـب الفنيـة لمنصـات التعلـم. فتهتـم معايـير (IMS QTI) بجـودة أدوات التقييـم الإلكـتروني وفعاليتهـا، في حـين تركـز معايـير (IEEE) في البنيـة التحتيـة التقنيـة. (IMS Global Learning Consortium, 2024; IEEE LTSC, 2024). وطـورت منظمتـا (ISO) و(IEC) معايـير شـاملة للتعلـم الإلكـتروني، تشـمل جوانـب مهمـة، مثـل: أمـن البيانـات وخصوصيتهـا، وجـودة المحتـوى التعليـمي، وآليـات التفاعـل والتواصـل، وقابليـة التوسـع والتطويـر. وتسـاعد هـذه المعايـير في بنـاء منصـات آمنـة وموثوقـة لتعليـم العربيـة لغـةً ثانيـةً (International Organization for Standardization, 2023).

وفي سياق آخر، قدم المركز الوطني للتعليم الإلكتروني إطاراً متكاملاً يشمل معايير جودة التعليم الإلكتروني، وإطار المهارات الرقمية، ومعايير تطوير المحتوى الإلكتروني، ومواصفات منصات التعلم. كما طور مكتب التربية العربي وثائق معيارية لتعليم اللغة العربية وتطوير المحتوى الرقمي العربي، تشمل المعايير المهنية للمعلمين ومعايير تقويم البرامج التعليمية (المركز الوطني للتعليم الإلكتروني، ٢٠٢٣).

وفي دراسة تقييمية مهمة، كشف القاضي (٢٠١٩) عن ضعف واضح في تطبيقات تعليم العربية لغة ثانية المتوافرة في متاجر التطبيقات. وقد طور نموذجاً تقييميًّا يتضمن ست فئات وثلاثين معيارًا، وأظهرت نتائج الدراسة أن التطبيقات المقيَّمة لم تحقق سوى ٥٤٪ من المعايير المطلوبة. وقدمت دراسة الغامدي وفلاته (٢٠٢٣) إسهامًا مهمًّا في مجال تصميم المنصات التعليمية لذوي اضطرابات النطق واللغة. وتوصلت الدراسة إلى تطوير قائمة معايير تشمل تسعة معايير رئيسة و١٠٤ مؤشرات، مع توصيات بتطوير نماذج تصميم تعليمية تراعي احتياجات هذه الفئة من المتعلمين.

تؤكد هذه المراجعة الشاملة للأدبيات أهمية تطوير إطار معياري متكامل يجمع بين المعايير التربوية والتقنية والتنظيمية، مع مراعاة خصوصية اللغة العربية واحتياجات المتعلمين المختلفة. كما تُبرز ضرورة الإفادة من المعايير العالمية مع تكييفها لتناسب سياق منصات تعليم اللغة وخصائص اللغة العربية.

## نطاق المعايير وحدودها

تشمل المعايير المقدمة في هذا الفصل أهم المبادئ والأسس التي يجب مراعاتها عند تخطيط وتصميم وبناء وتنفيذ المنصات الإلكترونية لتعليم اللغة العربية وتعلمها، وتتضمن كل مراحل تطوير المنصات التعليمية الإلكترونية، وقد بُنيت استنادا إلى نتائج تحليل ودراسة عدد من المعايير والوثائق ذات الصلة.

وصُنِّفت هذه المعايير إلى تسع مجموعات رئيسة، تندرج تحت كلٍّ منها مجموعة من المعايير الموزعة في فئات متعددة؛ بما يسهم في تسهيل تطبيقها وفقًا للسياق التعليمي المناسب.

## قائمة المعايير[1]:

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| **أولا: معايير إعداد المحتوى التعليمي الرقمي** | ١. التمهيد بتعليمات واضحة عن برنامج تعليم اللغة | أ. كتابة تعليمات واضحة للمستخدم تتضمن نبذة عن برنامج تعليم اللغة، وأهم المعلومات الخاصة به. | إنشاء نصوص وإرشادات موجهة للمستخدمين توضح لهم كيفية استخدام برنامج معين لتعلم اللغة. هذه التعليمات ينبغي أن تكون مفهومة وسهلة المتابعة، وتتضمن نبذة مختصرة عن البرنامج، أي: تقديم مقدمة تعريفية توضح أهداف البرنامج ومميزاته الرئيسة. | نسبة وضوح التعليمات للمستخدمين | استبيان رضا المستخدم |
| | | ب. توضيح مكونات برنامج تعليم اللغة، والجدول الزمني للتعلم. | يهدف إلى جعل المتعلم على دراية تامة بما سيتعلمه، ومتى سيتعلمه، وكيفية تنظيم عملية التعلم؛ مما يسهل عليه متابعة تقدمه وفهم مسار التعلم بوضوح. | نسبة وضوح مكونات البرنامج والجدول الزمني للتعلم | تدقيق المحتوى من خلال قائمة تحقق تشمل: وجود أهداف، محتوى، أنشطة، تقويم، وجدول زمني. |

(١) تم إدراج مؤشرات أداء متوسطة، لا تتطلب - غالبا - إجراءات مطولة، بحيث يمكن الإفادة منها بيسر وسهولة.

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ج. توضيح قنوات التواصل وطلب الدعم والمساندة لمستخدمي المنصة. | تقديم معلومات واضحة حول كيفية تواصل المستخدمين مع فريق الدعم الفني أو الأكاديمي للحصول على المساعدة أو حل المشكلات المتعلقة باستخدام المنصة التعليمية. | وجود طرق الدعم والمساندة | تدقيق المحتوى عن طريق قائمة تحقق. |
| | | د. كتابة سياسات ومنهج التعلم الخاص بالمنصة. | تحديد وتوثيق القواعد والمبادئ التعليمية التي تنظم عملية التعلم والمنهجية المتبعة في المنصة. | وجود السياسات | مراجعة السياسات عن طريق قائمة تحقق. |
| | | هـ. توضيح إجراءات التقدم في مراحل ومهمات البرنامج التعليمي. | شرح الخطوات المتسلسلة والمتطلبات اللازمة للانتقال بين مراحل البرنامج التعليمي ومهماته المختلفة. | نسبة المهام التي لها معايير انتقال واضحة | مراجعة دليل المقرر |
| | | و. تحديد المتطلبات والمعارف والمهارات الرئيسة التي قد يحتاجها المستخدم قبل البدء في تعلم البرنامج. | توضيح المهارات والمعارف الأساسية التي ينبغي أن يمتلكها المتعلم قبل البدء في البرنامج. | نسبة الدورات التي تعرض متطلبات مسبقة | فحص وصف المقررات عن طريق قائمة تحقق للمتطلبات والمهارات |
| | | ز. تحديد الخصائص التقنية للمنصة والأجهزة المتوافقة معها. | تقديم معلومات تفصيلية عن المواصفات التقنية للمنصة، ومدى دعمها للأجهزة والأدوات التقنية المختلفة. | نسبة الأجهزة/ المتصفحات المغطاة | اختبار تقني وقائمة تحقق بمدى توفر الخصائص المطلوبة. |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ح. التعريف بمقدم -أو مقدمي- البرنامج التعليمي إذا كان المحتوى معتمدًا على تقديم فرد أو مجموعة من المدربين. | تقديم نبذة تعريفية عن المدربين وخبراتهم ومؤهلاتهم في مجال المحتوى التعليمي. | نسبة المقررات التي تعرض ملفات تعريف المدربين | تدقيق ملفات التعريف |
| | ٢. استيفاء السمات الأساسية في صياغة أهداف البرنامج التعليمي ونواتج تعلمه | أ. الالتزام بالضوابط المنهجية في صياغة أهداف البرنامج ونواتج التعلم. | تحديد أهداف التعلم بوضوح وفقًا لمبادئ تعليمية معتمدة، بحيث تكون قابلة للقياس والتحقق. | نسبة أهداف البرنامج ونواتج التعلم التي تمت صياغتها بطريقة قابلة للقياس. | مراجعة محتوى أهداف البرنامج باستخدام قائمة تحقق |
| | | ب. اتساق أهداف البرنامج مع المعايير والأطر المحلية والعالمية لتعليم اللغة الثانية، وتعليم اللغة العربية لغة ثانية. | التأكد من أن أهداف تعلم اللغة تتناسب مع المعايير الدولية والمحلية المعتمدة لتعليم اللغة العربية لغة ثانية أو أجنبية، مثل: معايير الإطار الأوروبي المشترك لتعليم اللغات ووثائق تعليم العربية لغة ثانية. | نسبة المواءمة مع إطار معتمد | مراجعة خارجية |
| | | ج. توفير آلية لتتبع تقدم المتعلم في تحقيق أهداف ونواتج التعلم الخاصة بالبرنامج. | تقديم نظام يقيس ويراقب تقدم المتعلم في الوصول إلى أهداف التعلم المقررة، وتقديم المحتوى وفقًا لنظام تحليل ذكي لسلوك المستخدم وأدائه. | وجود آلية لكشف تقدم المتعلم | مراجعة نواتج القياس في المنصة |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | د. توزيع متوازن لنواتج التعلم على الوحدات التعليمية في البرنامج. | توزيع الأهداف التعليمية توزيعاً متوازناً عبر الوحدات المختلفة؛ لضمان تحقيق نواتج التعلم المطلوبة تدريجياً. | معامل تشتت توزيع النواتج | تحليل بيانات المقرر |
| | | هـ. ضمان الوصول المستمر للمتعلم أثناء التعلم إلى نواتج التعلم المستهدفة. | التأكد من وصول المتعلم إلى ملخصات الأهداف التعليمية في أي وقت أثناء البرنامج؛ لتعزيز فهمه ومستوى تقدمه. | نسبة نجاح المتعلم في الاختبار البعدي | مراجعة نتائج الاختبار القبلي والبعدي |
| | | و. توافق نواتج التعلم مع الفئة المستهدفة والمستوى الأكاديمي للمستخدمين. | تصميم الأهداف التعليمية بما يتناسب مع قدرات الفئة المستهدفة، واحتياجاتها، ومستواها الأكاديمي. | نسبة الوحدات المناسبة للمستوى | استبيانات تقييم المستخدمين (الطلاب أو المتدربين). |
| | ٣. الالتزام بخصائص التقويم الجيد للتعلم | أ. قياس مستوى إنجاز المتعلم وربطه بتحقيق النواتج التعليمية. | تقييم مدى تقدم المتعلم بناءً على مدى تحقيقه للأهداف التعليمية المحددة. | نسبة المتعلمين الذين حققوا نواتج التعلم المستهدفة. | جمع وتحليل بيانات التقييم والأداء (نتائج الاختبارات، التقييمات العملية، المشاريع، العروض). |
| | | ب. توضيح آلية وتقنيات التقويم والقياس المعتمدة في البرنامج التعليمي. | شرح الأدوات والأساليب المستخدمة لتقييم أداء المتعلمين وقياس تقدمهم في البرنامج. | نسبة المقررات ذات سياسة منشورة | تدقيق السياسات |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ج. تقديم تعليق وتصحيح فوري وآلي للمتعلم في الاختبارات الإلكترونية الموضوعية. | تقديم ملحوظات وتصحيحات فورية بعد إتمام المتعلم للاختبارات الإلكترونية؛ لتمكينه من التعلم الفوري من أخطائه. | توفر تصحيح فوري | مراجعة منصة التعلم |
| | | د. وضع آلية معتمدة للتقويم الذاتي لمشاريع و مهمات وأنشطة المتعلمين. | إنشاء أدوات ومعايير تمكن المتعلمين من التقييم الذاتي لأعمالهم ومشاريعهم. | نسبة المهام مع آلية التقويم الذاتي للمتعلم | مراجعة الأنشطة |
| | | هـ. توفير أدوات تساعد المتعلمين على تقييم أدائهم، وتحديد مدى تقدمهم في تحقيق نواتج التعلم. | إتاحة فرص التقييم الذاتي للمستخدمين، وتمكينهم من قياس مدى تقدمهم في البرنامج التعليمي. | نسبة المقررات أو الأنشطة التي تتضمن أدوات تقويم ذاتي للمتعلمين. | مراجعة منصة التعلم |
| | ٤. توافر الموارد الإلكترونية الحديثة والداعمة | أ. تقديم مجموعة متنوعة من الموارد الإلكترونية لتعليم اللغة باستخدام أحدث التقنيات التعليمية. | تتيح المنصة مواد تعليمية رقمية متنوعة، مثل: مواد تعليم مهارات اللغة، والمعاجم التعليمية، والأنشطة التفاعلية، وغيرها، وتعتمد على أحدث التقنيات المتاحة لتسهيل تعلم اللغة. | نسبة الموارد الحديثة لآخر ١٢ شهرا | مراجعة محتوى المقررات أو المنصة التعليمية |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ب.توظيف الموارد التعليمية الرقمية في أداء الأنشطة والمهمات التعليمية وعمليات القياس والتقويم. | تستخدم المنصة الموارد الإلكترونية في تنفيذ الأنشطة التعليمية وتقييم أداء المتعلمين. | نسبة الأنشطة التي تستخدم موارد رقمية | مراجعة الأنشطة |
| | | ج. ربط المستخدم بمصادر المعلومات والمراجع الأصيلة والموثوقة. | توجه المنصةُ المستخدمَ إلى مصادر موثوقة وأصلية، وتراعي حقوق الملكية الفكرية في جميع ما تعرضه عليه من مواد تعليمية مختلفة. | نسبة المصادر ذات الأصيلة والموثوقة | تدقيق المصادر |
| | | د. الالتزام بحقوق الملكية الفكرية للمواد التعليمية. | | نسبة انضباط المواد بحقوق الملكية الفكرية | تدقيق المواد |
| | | هـ. تصميم مواد تعليم اللغة العربية الرقمية وفق أحدث وأفضل نظريات وممارسات تعليم اللغة الثانية وتعلمها. | إعداد المواد التعليمية الرقمية بناءً على أحدث النظريات والطرق المثبتة في تعليم اللغات، مثل: المدخل التواصلي، والمنهج التكاملي في تعليم اللغة. | نسبة المواد التعليمية الرقمية المصممة وفق معايير تعليم اللغة الثانية الحديثة. | مراجعة دورية |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| **ثانيًا: تجـربة المتعلم** | ١. تصـميـم المهمـات والأنشطـة بطـريـقـة تساعـد على التـفاعل والتـعـلـم النشط | أ. تصميم مهمات وأنشطة لغوية إلكترونية تساعد المتعلم على تحقيق النواتج المستهدفة لتعليم اللغة العربية لغة ثانية. | تصميم الأنشطة والمهام اللغوية الإلكترونية بطريقة منهجية تهدف إلى دعم المتعلم في تحقيق الأهداف التعليمية المحددة. تركز هذه الأنشطة في تطوير المهارات اللغوية الأربع (القراءة، والكتابة، والاستماع، والتحدث) وتتناسب مع مستوى المتعلم؛ مما يعزز تقدمه التدريجي في اكتساب اللغة. يشمل التصميم أنشطة متنوعة، مثل: تمارين القواعد، والاختبارات التفاعلية، والمحادثات الافتراضية التي تتيح للمتعلمين تطبيق ما تعلموه في سياقات عملية؛ مما يساعدهم على استيعاب اللغة بسرعة وفعالية. | نسبة الأنشطة المرتبطة بنواتج تعلم واضحة | مراجعة محتوى المنصة والأنشطة |
| | | ب. تقديم المهمات والأنشطة التعليمية التي تعزز فرص التعلم النشط، وتشجع تفاعل المتعلمين. | تشجيع المتعلمين على المشاركة في أنشطة تفاعلية، مثل: المحادثات الحية، والمشاريع الجماعية، والألعاب التعليمية التي تتطلب التفكير النقدي واستخدام المهارات اللغوية في مواقف حقيقية. هذا النوع من التعلم يعزز مشاركة المتعلمين ويدعم تعلمهم من خلال التجربة العملية؛ مما يؤدي إلى نتائج تعليمية أكثر عمقاً وديمومة. | نسبة الدروس التي تتضمن أنشطة نشطة (مشاريع، محادثات...) | مراجعة محتوى المنصة والأنشطة |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ج. ضمان التفاعل المستمر بطرق متنوعة بين المتعلم والمعلم، أو بين المتعلم والمواد التعليمية. | تصميم الأنشطة والمهمات لتوفير فرص تفاعل مستمر بين المتعلم والمعلم أو بين المتعلم والمواد التعليمية. ويتم ذلك عبر أدوات تفاعلية، مثل: المحادثات الفورية، والمنتديات الإلكترونية، والتغذية الراجعة الفورية التي تسهل التواصل الفعال والدعم المستمر. هذا التفاعل المتواصل يساعد المتعلم على الفهم العميق للمحتوى والتمكن الجيد من اللغة، إضافة إلى معالجة أي صعوبات قد يواجهها خلال عملية التعلم. | نسبة المواد ذات الأنشطة التفاعلية | مراجعة محتوى المنصة والأنشطة |
| | ٢. توفير بيئة تعليمية إلكترونية داعمة للتعلم. | أ. توضيح سياسات التواصل والتفاعل بين مكونات العملية التعليمية: (المعلم - المتعلم - المواد التعليمية). | تتضمن المنصة آليات واضحة لشرح كيفية التواصل والتفاعل بين جميع الأطراف المعنية في العملية التعليمية، سواء بين المعلم والمتعلم، أم بين المتعلم والمحتوى التعليمي. وتشمل هذه الآليات: الرسائل المباشرة، والمنتديات، وجلسات النقاش عبر الفيديو. هذه السياسات تهدف إلى تسهيل الحوار المستمر، وتقديم الدعم الفوري، وضمان بيئة تعليمية تفاعلية ومنظمة؛ مما يعزز تجربة التعلم، ويضمن تحقيق الأهداف التعليمية بفعالية أكبر. | نسبة توفر سياسات تواصل واضحة | مراجعة وثائق السياسات |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ب. تهيئة بنية تقنية آمنة ومتكاملة لإدارة محتوى المنصة التعليمية، وفقًا لأحدث تقنيات الويب والتقنيات التعليمية. | تعتمد المنصة على بنية تقنية حديثة وآمنة تضمن إدارة فعالة للمحتوى التعليمي، باستخدام أحدث تقنيات الويب لضمان استقرار وأمان المنصة، مع حماية بيانات المستخدمين وضمان سرية المعلومات. هذه البنية التقنية تتيح الوصول السلس إلى المحتوى التعليمي وتدعم استخدام أدوات تعليمية متطورة، مثل: أنظمة إدارة التعلم (LMS)؛ مما يتيح بيئة تعليمية موثوقة ومتكاملة تدعم تفاعل المتعلمين ووصولهم إلى المعلومات بسهولة وأمان. | عدد معايير الأمان والتقنيات الحديثة المطبقة | تدقيق في للبنية التقنية |
| | | ج. دعم البيئة التقنية للمنصة نماذج متنوعة في التعلم الإلكتروني والتعلم عن بعد (التعلم المتزامن - غير المتزامن - المدمج - الافتراضي - التعلم باللعب - التعلم المبني على المهمات أو المشروعات). | تقدم المنصة دعمًا شاملاً لمجموعة متنوعة من نماذج التعلم الإلكتروني والتعلم عن بعد؛ مما يتيح للمتعلمين مرونة في اختيار النمط الذي يناسبهم. تشمل هذه النماذج: التعلم المتزامن (مثل الفصول الافتراضية الحية)، والتعلم غير المتزامن (مثل المحاضرات المسجلة)، والتعلم المدمج (مزيج من التعلم التقليدي والإلكتروني)، والتعلم الافتراضي، والتعلم باللعب، والتعلم القائم على المشاريع أو المهمات. هذا التنوع في الأساليب التعليمية يتيح للمتعلمين تجربة تعليمية مخصصة تلبي احتياجاتهم الشخصية وتفضيلاتهم؛ مما يدعم تحقيق الأهداف التعليمية بفعالة. | نسبة النماذج التعليمية المتاحة | مراجعة خصائص المنصة |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| **ثالثًا: التقنيات التعليمية والبيانات وسمات الدمج والمشاركة** | ١. توفير الأدوات والتقنيات التي تدعم مناهج تعليم اللغة الثانية وتعلمها | أ. توظيف المنصة لأحدث الأدوات والتقنيات التعليمية لتحقيق أهداف تعلم اللغة الثانية وفقًا لأفضل النظريات والممارسات التطبيقية في ميدان التعليم الإلكتروني وتعليم اللغة الثانية وتعلمها. | تعتمد المنصة على أحدث الابتكارات في مجال التعليم الإلكتروني وتعليم اللغات، وتشمل: التعلم التفاعلي، والذكاء الاصطناعي؛ لضمان تحقيق أهداف تعلم اللغة الثانية بفعالية، وتَوافُق مع أحدث نظريات اللسانيات التطبيقية. | نسبة الأدوات المحدثة المستخدمة في المنصة | مراجعة خصائص المنصة + التدقيق الفني |
| | | ب. تمكين المتعلمين -عبر التقنيات المختلفة- من تطبيق المنهج التكاملي في تعلم مهارات اللغة الثانية وعناصرها. | تُقدِّم التقنيات التعليمية أدوات متعددة تمكن المتعلمين من دمج المهارات اللغوية المختلفة (القراءة، والكتابة، والاستماع، والتحدث) في عملية تعلم متكاملة؛ مما يساعد المتعلمين على تحسين فهمهم للغة واستخدامها بتوازن. | نسبة الأدوات التي تتيح تطبيق المنهج التكاملي | تحليل الأدوات التعليمية |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ج. إتاحة الأدوات والتقنيات التعليمية فرصاً لممارسة التعلم النشط والتفاعل والتواصل المستمر بين المعلم والمتعلمين والمناهج والمواد اللغوية التعليمية. | تسهم الأدوات التعليمية في تعزيز التفاعل المستمر بين المعلمين والمتعلمين عن طريق الأنشطة التفاعلية والاتصالات الفورية؛ مما يدعم التعلم النشط ويعزز التواصل بين جميع عناصر العملية التعليمية. | نسبة الأنشطة التفاعلية المدعومة | تحليل الأدوات التعليمية |
| | ٢. إتاحة المنصة خواص الدمج والمشاركة | أ. تمكين مستخدمي المنصة من مشاركة تجربتهم وتيسير تداول إنجازات تعلم اللغة في الشبكات الاجتماعية. | تتيح المنصة مشاركة إنجازات المتعلمين على وسائل التواصل الاجتماعي؛ مما يعزز تجربة التعلم ويمكن المتعلمين من عرض تقدمهم ومشاركة خبراتهم مع الآخرين. | نسبة المستخدمين الذين يشاركون إنجازاتهم | تتبع بيانات المشاركة على المنصة |
| | | ب. إتاحة إمكانية دمج أدوات المنصة وبياناتها في منصات أخرى عن طريق تفعيل تقنية (API) لتبادل البيانات عبر الشبكة. | تدعم المنصة تكامل الأدوات والبيانات مع منصات أخرى عن طريق واجهات برمجة التطبيقات (API)؛ مما يسهل نقل البيانات وتكامل الخدمات التعليمية عبر الشبكات المختلفة. | وجود واجهة برمجة التطبيقات (API) مفعلة وقابلة للاستخدام. | فحص المنصة |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| **رابعًا: دمج تقنيات الذكاء الاصطناعي** | ١. توفير أدوات تساعد على التنبؤ بأداء المتعلم | أ. تحليل أنماط الأداء السابق للمتعلمين. | يدرس النظام تاريخ تعلم المستخدم بعمق، متتبعاً كل خطواته وتفاعلاته. يشمل التحليل معدل إكمال التمارين، ودقة الإجابات، وسرعة الاستجابة، والوقت المستغرق في كل نشاط. هذه البيانات تشكل صورة شاملة عن نمط تعلم المستخدم وتطوره عبر الزمن؛ مما يساعد في فهم نقاط القوة والضعف لديه. | عدد تقارير تحليل الأداء الصادرة شهريًا | مراجعة تقارير النظام |
| | | ب. التنبؤ بالصعوبات المحتملة التي قد يواجهها المتعلم. | باستخدام نماذج التعلم الآلي، يستطيع النظام توقع المجالات التي قد يواجه فيها المتعلم صعوبات مستقبلية. ويستند هذا التنبؤ إلى تحليل أنماط الأخطاء السابقة، ومقارنتها مع بيانات المتعلمين الآخرين؛ مما يتيح التدخل المبكر قبل تفاقم المشكلات. | وجود قائمة بالصعوبات المحتملة التي قد تواجه المتعلم ومدى شمولها في وثائق البرنامج. | مراجعة محتوى البرنامج والوثائق. |
| | | ج. تحديد احتمالات التسرب والانقطاع عن الدراسة. | يراقب النظام مؤشرات -مثل: تراجع معدل المشاركة، أو زيادة الأخطاء، أو تغير نمط الاستخدام- حتى تساعد في التنبؤ باحتمال انقطاع المتعلم عن التطبيق؛ مما يتيح اتخاذ إجراءات وقائية لضمان استمراره في التعلم. | وجود وثائق أو تقارير تحدد احتمالات تسرب وانقطاع المتعلمين عن الدراسة في البرنامج. | مراجعة وثائق البرنامج . |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | د. اقتراح تدخلات استباقية لتكييف المحتوى التعليمي مع احتياجات المتعلم. | بناءً على التحليلات السابقة، يقدم النظام توصيات محددة للتدخل المبكر. قد تشمل هذه التدخلات: تعديل مستوى الصعوبة، أو تقديم تمارين إضافية، أو توفير دعم إضافي في مجالات معينة قبل ظهور المشكلات. | عدد التدخلات الاستباقية المقترحة | مراجعة خطط التدخل في النظام |
| | | هـ. تخصيص خطط التعلم المستقبلية بناء على تحليل أداء المتعلم. | يصمم النظام مسارات تعلم مخصصة تتناسب مع أهداف المتعلم وقدراته. تأخذ هذه الخطط في الاعتبار سرعة التعلم المفضلة، وأوقات التعلم المثالية، والموضوعات التي تحتاج إلى تركيز أكبر. | نسبة الخطط الفردية المخصصة | مراجعة ملفات خطة التعلم |
| | ٢. توفير أدوات لتحليل أنماط التعلم المختلفة | أ. رصد وقت التعلم الأمثل بناء على تحليل بيانات المتعلم وأدائه. | يحلل النظام أنماط استخدام المتعلم لتحديد الأوقات التي يظهر فيها أفضل أداء. ويراعي هذا التحليل عوامل عدة، مثل: الوقت، واليوم، ومدة الجلسات، وفترات الراحة المثلى؛ مما يساعد في اقتراح جدول تعلم أكثر فعالية. | عدد التوصيات الزمنية للمتعلمين | مراجعة تقارير التحليل |
| | | ب. تحديد المواد التعليمية الأكثر فعالية عن طريق تحليل بيانات المتعلمين. | يقيّم النظام فعالية مختلف أنواع المواد التعليمية (فيديوهات، أو تمارين تفاعلية، أو ألعاب تعليمية) بالنسبة للمتعلم المحدد، ويحدد أي المواد تؤدي إلى أفضل نتائج التعلم. | قائمة تصاعدية للمواد الأكثر فعالية | مراجعة النظام |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ج. تحليل إستراتيجيات التعلم الناجحة للتركيز في الأكثر فائدة للمتعلم. | يدرس النظام أنماط التعلم التي أدت إلى أفضل النتائج للمتعلم، مثل التسلسل المعين للتمارين، أو مزيج محدد من الأنشطة. ويستخدم هذه المعلومات لتحسين عملية التعلم. | قائمة تصاعدية للاستراتيجيات الأكثر فعالية | مراجعة النظام |
| | | د. تتبع تفضيلات المتعلم واقتراح المحتوى التعليمي المناسب له بناء على ذلك. | يراقب النظام اختيارات المتعلم وتفضيلاته، مثل: أنواع التمارين المفضلة، والموضوعات التي يقضي فيها وقتًا أطول، وأسلوب التعلم الذي يناسبه. | نسبة التوصيات المخصصة حسب التفضيلات | تحليل بيانات تفضيلات المستخدم |
| | | هـ. تحسين مسارات التعلم بناء على نتائج تحليل بيانات الاستخدام. | يعدل النظام ويحسن مسار التعلم باستمرار بناءً على التقدم والنتائج. يشمل ذلك: تعديل مستوى الصعوبة، وإضافة أو إزالة مواد تعليمية، وتغيير ترتيب الدروس. | نسبة المسارات التعليمية التي تم تعديلها بناءً على نتائج تحليل بيانات الاستخدام. | مراجعة النظام |
| | ٣. تقديم توصيات للمحتوى بناء على تقنيات تحليل بيانات سلوك المتعلم | أ. اقتراح موارد تعليمية مناسبة لمتعلم اللغة ومستواه مبنية على تحليل البيانات. | يقدم النظام توصيات بموارد إضافية تتناسب مع مستوى المتعلم واهتماماته، مثل: مقاطع الفيديو، أو المقالات، أو التمارين الإضافية التي تدعم تعلمه. | التحقق من توفر الموارد التعليمية الموصى بها لكل متعلم بناءً على تحليل بيانات أدائه. | مراجعة النظام |
| | | ب. تحديد المواد التكميلية المناسبة لمتعلم اللغة. | يقترح النظام مواد تكميلية تساعد في تعزيز فهم متعلم اللغة للموضوعات الصعبة، أو توسيع معرفته في المجالات التي يبدي فيها اهتماماً خاصًا. | نسبة المواد التكميلية المقترحة | مراجعة النظام |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ج. التوصية بأنشطة تعزيزية مناسبة لأداء المتعلم. | يقدم النظام اقتراحات لأنشطة إضافية تساعد في تثبيت المعرفة وتطبيقها، مثل: التمارين العملية، أو المشاريع الصغيرة، أو التحديات اللغوية. | عدد الأنشطة التعزيزية الموصى بها | مراجعة النظام |
| | | د. اقتراح مجموعات تعلم مناسبة وفقًا لتحليل سلوك المتعلمين وتفضيلاتهم. | يحدد النظام مجموعات التعلم التي تناسب مستوى المتعلم واهتماماته؛ مما يتيح له التفاعل مع متعلمين آخرين لديهم أهداف وتحديات مماثلة. | نسبة المتعلمين ضمن مجموعات مخصصة | تقارير تكوين المجموعات |
| | | هـ. تقديم محتوى مخصص حسب اهتمامات المتعلمين. | يخصص النظام المحتوى التعليمي ليتناسب مع اهتمامات المتعلم الشخصية؛ مما يجعل عملية التعلم أكثر جاذبية وارتباطاً بحياته اليومية. | نسبة المحتوى المخصص وفق اهتمامات المتعلين. | مراجعة النظام |
| | ٤. توظيف الذكاء الاصطناعي في دعم وتخصيص تجربة متعلم اللغة | أ. توظيف المنصة تقنيات الذكاء الاصطناعي لتحسين تعليم مهارات اللغة وعناصرها. | تستخدم المنصة تقنيات الذكاء الاصطناعي لتقديم تعليم مخصص للمهارات اللغوية المختلفة، مثل تحسين مهارات النطق أو الكتابة، عن طريق أنظمة ذكية تفاعلية تساعد على تصحيح الأخطاء وتقديم ملاحظات فورية. | نسبة المهارات اللغوية المدعومة بتقنيات الذكاء الصناعي | مراجعة خصائص المنصة |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ب. توظيف نماذج الذكاء الاصطناعي التوليدي في إنتاج المواد التعليمية. | تعتمد المنصة على نماذج الذكاء الاصطناعي التوليدي لإنشاء مواد تعليمية مبتكرة، مثل: التمارين اللغوية أو الاختبارات المصممة خصيصى لاحتياجات المتعلمين؛ مما يُقدّم محتوى ديناميكيًا ومخصصًا. | عدد المواد المنتجة بالذكاء الصناعي | مراجعة خبير للمواد المولدة |
| | | ج. استثمار تقنيات الذكاء الاصطناعي في تخصيص تجربة متعلم اللغة. | استخدام الذكاء الاصطناعي لتخصيص تجربة التعلم لكل فرد، بتعديل المحتوى والأنشطة بحسب مستوى المتعلم وأسلوب تعلمه؛ مما يساعد على تحسين تجربة التعلم وزيادة فعالية التعليم. | نسبة الأنشطة المخصصة بالذكاء الاصطناعي | مراجعة نظام التخصيص |
| خامسًا: دعم متعلمي اللغة ومعلميها | ١. مراعاة الاحتياجات الخاصة لفئات المستخدمين | أ. تحقيق معايير الوصول إلى المحتوى عبر الإنترنت وفقًا لما ورد في الدليل المرجعي لمعايير الوصول إلى محتوى الويب لمختلف فئات المستخدمين، ولا سيما من ذوي الاحتياجات الخاصة (Web Content Accessibility Guidelines - WCAG) | يشير هذا المعيار إلى ضرورة تصميم المنصة التعليمية، بحيث تلتزم بمعايير الوصول العالمية (WCAG) التي تضمن أن المحتوى متاح للجميع، بما في ذلك ذوو الاحتياجات الخاصة، من خلال تحسين إمكانية الوصول عبر النصوص البديلة للصور، وأدوات التنقل الملائمة، والتوافق مع التقنيات المساعدة مثل قارئات الشاشة. | نسبة الالتزام بمعايير الوصول WCAG | تدقيق الوصول Accessibility Audit |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ب. إتاحة المحتوى التعليمي المناسب لاحتياجات الفئات المستفيدة. | تضمن المنصة تصميم المحتوى التعليمي بما يلبي احتياجات مختلف المتعلمين، بما في ذلك ذوي الاحتياجات الخاصة، من خلال تقديم مواد تعليمية قابلة للتكيف مع احتياجاتهم وتفضيلاتهم التعليمية. | نسبة المواد التعليمية المصممة لتلبية احتياجات الفئات المستفيدة. | مراجعة المحتوى التعليمي |
| | | ج. توضيح سياسات وأنظمة الدعم الفني المتكامل لمستعملي المنصة. | يجب التوضيح الكامل لسياسات وآليات الدعم الفني؛ ليتمكن المستخدمون من الوصول إلى الدعم الفني لأي مشكلات تقنية أو استفسارات تتعلق باستخدام المنصة، عبر قنوات متعددة، مثل: البريد الإلكتروني، أو الدردشة المباشرة، أو الهاتف. | عدد قنوات الدعم الفني المتكامل المتاحة لمستخدمي المنصة. | مراجعة وثائق الدعم الفني |
| | ٢. تصميم واجهات المستخدم بطريقة تراعي تيسير عمليات التنقل والوصول والتخصيص | أ. توفير أدلة إرشادية أو خرائط تفاعلية تمكن المستخدم من فهم بنية المنصة، وكيفية التنقل بين مكوناتها. | تُصمَّم واجهة المستخدم تصميما واضحاً ومنظماً يمكن المتعلمين والمعلمين من التنقل بسهولة بين مختلف أقسام المنصة، وتُقدَّم المساعدة الكافية للمستخدمين للتعرف على خارطة المنصة وطرق التنقل بين مكوناتها، من خلال مجموعة من الأدلة الإرشادية المتنوعة. | نسبة المنصات المزودة بأدلة وخرائط تفاعلية | مراجعة واجهة الاستخدام |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ب. تيسير الوصول إلى المحتوى التعليمي، والأخذ في الاعتبار مراعاة الوصول لمختلف فئات متعلمي اللغة. | تصميم المحتوى التعليمي ليكون سهل الوصول لمختلف المتعلمين، بما في ذلك متعلمي اللغة من المستويات المختلفة أو من ذوي الاحتياجات الخاصة؛ مما يتطلب استخدام أدوات سهلة الاستخدام وتقديم المحتوى بطرق متنوعة لتلبية احتياجات الجميع. | نسبة المحتوى المتاح لمختلف الفئات | تحليل المحتوى |
| | | ج. تمكين المستخدم من تخصيص تجربة التعلم عن طريق إتاحة عدد من خيارات التحكم في عرض المحتوى، مثل: (تكبير النص أو تصغيره، وتقنيات قراءة النص وكتابته آليًا). | تتيح المنصة خيارات تخصيص متعددة، مثل: تكبير أو تصغير النص، أو تغيير الألوان، أو استخدام أدوات قراءة النصوص تلقائيًا؛ مما يسمح للمستخدمين بتعديل تجربة التعلم وفقًا لاحتياجاتهم الخاصة. | نسبة خيارات التخصيص المتاحة لعرض المحتوى. | مراجعة إعدادات النظام |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | د. توظيف النماذج والأمثلة لتوضيح كيفية التفاعل مع الأنشطة والمهمات وإجراءات التقدم في المستويات اللغوية التي يتضمنها البرنامج. | شرح كيفية تقدم المتعلمين عبر مستويات اللغة المختلفة بوضوح، مع تقديم نماذج عملية تشرح كيفية التفاعل مع الأنشطة والمهمات التعليمية؛ مما يساعد المتعلمين على فهم مسار التعلم وتحقيق الأهداف اللغوية. | نسبة الوحدات المزوّدة بنماذج توضيحية | مراجعة الموارد التعليمية |
| | ٣. تحقيق بيئة إلكترونية تفاعلية وفعالة | أ. مناسبة استعمال الصور والوسائط المتعددة لمحتوى المنصة التعليمي. | تُختار الصور والوسائط المتعددة بما يتوافق مع المحتوى التعليمي ويخدم أهداف التعلم، مع توفير نصوص بديلة ووصف للوسائط لضمان فهم جميع المتعلمين للمحتوى. | نسبة الوسائط المرتبطة بالأهداف التعليمية | تحليل محتوى الوسائط عن طريق قائمة تحقق. |
| | | ب. تيسير استعمال وعرض الوسائط المتعددة وتكييفها للعرض في مختلف أنواع الأجهزة الإلكترونية: (الحاسوب المكتبي - المحمول - الجهاز اللوحي الهاتف المحمول). | تصميم الوسائط المتعددة لتكون مرنة وقابلة للتكيف مع مختلف الأجهزة الإلكترونية؛ مما يتيح تجربة تعليمية سلسة، سواء عبر الحاسوب المكتبي، أم الأجهزة اللوحية، أم الهواتف المحمولة. | نسبة توافق الوسائط مع الأجهزة | اختبار الوسائط على الأجهزة |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ج. تقديم التغذية الراجعة الفورية بطرق متعددة للمتعلم في أثناء أداء المهمات والأنشطة المختلفة. | تقدم المنصة تغذية راجعة فورية ومتعددة الأشكال (نصية، وصوتية، ومرئية) عند إتمام المتعلم للمهام أو الأنشطة التعليمية؛ مما يساعده على فهم أخطائه وتحسين أدائه. | نسبة الأنشطة المزودة بتغذية راجعة فورية | مراجعة خصائص المنصة |
| | | د. تهيئة مجتمعات افتراضية للتواصل وتبادل الخبرات بين المتعلمين والمعلمين. | إنشاء مجتمعات افتراضية أو منصات تفاعلية داخل المنصة تتيح لمتعلمي اللغة ومعلميها التواصل وتبادل الأفكار والخبرات؛ مما يعزز التفاعل والتعلم التعاوني فيما بينهم. | عدد المجتمعات الافتراضية النشطة | مراجعة أنظمة المنتديات / المجموعات |
| **سادسًا: ضمان جودة برامج تعليم اللغة** | ١. التوافق مع معايير وأطر تعلم اللغة الثانية وتعليمها | أ. التوافق في إعداد البرامج اللغوية مع معايير الإطار الأوروبي المشترك لتعليم اللغات (CEFR)، أو الأطر المحلية والعالمية المماثلة لاعتماد برامج تعليم اللغة العربية لغة ثانية. | تتوافق البرامج اللغوية مع معايير معترف بها دوليًا ومحليًا، مثل الإطار الأوروبي المشترك لتعليم اللغات (CEFR)، لضمان تقديم تعليم لغة منظم وفعال يتوافق مع مستويات الكفاءة اللغوية المختلفة؛ مما يتيح مقارنة نتائج التعلم مع المعايير العالمية. | نسبة البرامج المتوافقة مع CEFR أو أطر مماثلة | مراجعة خطط البرامج ومقارنتها بالمعايير |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ب. تحقيق البرنامج أو الدورة اللغوية معايير برامج التعليم الإلكتروني والتعليم عن بعد المحلية والعالمية. | يتطلب البرنامج أو الدورة اللغوية الالتزام بمعايير التعليم الإلكتروني والتعليم عن بعد المعترف بها عالميًا ومحليًا، مثل: لوائح ومعايير التعليم الإلكتروني في المملكة العربية السعودية، لضمان الاعتماد وجودة المحتوى وأساليب التدريس المستخدمة في بيئة التعلم الرقمية. | نسبة الدورات المطابقة لمعايير التعليم الإلكتروني | تدقيق اعتماد البرامج |
| | ٢. التكامل والاتساق بين جميع عناصر مناهج تعليم اللغة الثانية في تقديم تجربة تعليمية ثرية | أ. تحقيق البرنامج أو الدورة اللغوية أهدافًا تعليمية واضحة مرتبطة باكتساب مهارات اللغة العربية وعناصرها وثقافتها. | يكون للبرنامج أهداف واضحة ومحددة ترتبط بتطوير مهارات اللغة العربية (الاستماع، والقراءة، والكتابة، والتحدث) وفهم ثقافتها؛ مما يساعد المتعلمين على الاكتساب الشامل والمتكامل للغة. | نسبة وضوح الأهداف في البرامج | مراجعة وثائق الأهداف |
| | | ب. مناسبة المحتوى التعليمي الإلكتروني لأحدث وأفضل مناهج وطرق تعليم اللغة الثانية وتعلمها. | تصميم المحتوى التعليمي الإلكتروني وفق أحدث النظريات والممارسات في تعليم اللغة الثانية، مثل: المنهج التواصلي والتكاملي في تعليم مهارات وعناصر اللغة الثانية؛ لضمان فعالية عملية التعلم وتوافقها مع الطرق الحديثة في تعليم اللغات. | نسبة توافق المحتوى مع المناهج الحديثة | تحليل محتوى المناهج |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ج. توافر المصادر والمواد التعليمية الإلكترونية المتوافقة مع مناهج وطرق التعلم المعتمدة في المنصة. | تتوافر مجموعة متنوعة من المصادر والمواد التعليمية الإلكترونية التي تتناسب مع المناهج وطرق التعلم المعتمدة في المنصة؛ مما يساعد المتعلمين على تعزيز فهمهم للمحتوى التعليمي. | نسبة توفر المصادر التعليمية الموافقة | مراجعة المكتبة التعليمية |
| | | د. تطبيق أساليب وإستراتيجيات تعليم فعالة تلبي احتياجات المتعلمين، وتتسق مع الأهداف ونواتج التعلم المستهدفة. | يمتلك المعلمون المشاركون في برامج التعليم المتزامن المعرفة والخبرة اللازمة لتعليم اللغة العربية لغةً ثانيةً، ويوظفونها بفعالية في مختلف مواقف التعلم. | نسبة استخدام استراتيجيات تدريس فعالة | متابعة أنشطة التدريس |
| | | هـ. المراجعة والتحديث المستمر للمحتوى التعليمي؛ للتأكد من تلبيته لاحتياجات متعلمي اللغة المختلفة. | تحديث المحتوى التعليمي بانتظام بناءً على أحدث التطورات في تعليم اللغة، مع مراعاة احتياجات المتعلمين المختلفة لضمان تقديم تعليم مخصص وفعال. | عدد التحديثات المنفذة سنويًا | مراجعة سجلات تحديث المحتوى |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| **سابعًا: تقييم إعداد برامج تعليم اللغة** | ١. التحقق من استيفاء الضوابط الأساسية لبناء البرامج اللغوية الإلكترونية | أ. دراسة خصائص واحتياجات الفئة المستهدفة من المتعلمين لضمان تقديم محتوى يتلاءم مع مستواهم اللغوي وخلفياتهم الثقافية. | مراعاة احتياجات الفئة المستهدفة ودراسة حاجاتها وخصائصها اللغوية؛ مما يعزز من فعالية التعلم. | نسبة البرامج التي تضمنت دراسة للفئة المستهدفة | مراجعة وثائق تحليل الفئة المستهدفة |
| | | ب. موافقة مكونات البرنامج لأحدث وأفضل نظريات وممارسات تعليم وتعلم اللغة الثانية (الخطط - المحتوى - الأنشطة - المعايير - طرق التعليم والتعلم - القياس والتقويم). | توافق جميع مكونات البرنامج، مثل: (الخطط، والمحتوى، والأنشطة، ومعايير التقييم) مع أفضل الممارسات والنظريات الحديثة في تعليم وتعلم اللغة الثانية؛ لضمان تحقيق أهداف التعلم بفعالية. | نسبة المكونات المتوافقة مع الممارسات الحديثة | تحليل البرامج ومقارنتها بالمعايير |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ج. التحقق من امتلاك المعلمين ومعدّي المحتوى التعليمي الإلكتروني الخبرات والمعارف والمهارات الملائمة. | التأكد من امتلاك المعلمين ومعدّي المحتوى التعليمي المهارات والخبرة المناسبة لتقديم تعليم عالي الجودة، بما يتناسب مع احتياجات المتعلمين وأهداف البرنامج والبيئة الإلكترونية للمنصة. | نسبة المعلمين / المعدين المؤهلين | مراجعة ملفات السيرة الذاتية والشهادات |
| | | د. كفاءة أدوات وطرق تقييم وقياس أداء متعلمي اللغة و معلميها إلكترونيًا. | تكون أدوات التقييم والقياس المستخدمة لتحديد أداء المتعلمين والمعلمين فعالة ودقيقة، معتمدة على أساليب حديثة في التقييم الإلكتروني؛ لضمان متابعة دقيقة للتقدم. | نسبة الأدوات الإلكترونية الفعّالة | تجربة أدوات التقييم وتحليل نتائجها |
| | | هـ. الاتساق والتناسب بين أهداف البرنامج أو الدورة اللغوية وجميع مكوناتها، مثل: (المحتوى التعليمي، والمهمات والأنشطة التعليمية، والقياس والتقويم، وطرق التعليم والتعليم). | يكون التناسق واضحاً بين أهداف البرنامج أو الدورة وجميع مكوناتها، بما في ذلك المحتوى التعليمي والأنشطة وطرق التقييم؛ لضمان تحقيق الأهداف التعليمية بمنهجية متكاملة. | نسبة التوافق بين أهداف البرنامج والمحتوى التعليمي. | مراجعة مصفوفة مواءمة الأهداف مع المكونات |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | ٢. توفير الوثائق المرجعية اللازمة | أ. توافر وثيقة مرجعية شاملة لإعداد البرنامج اللغوي ومحتواه التعليمي الإلكتروني. | إعداد وثيقة مرجعية شاملة توضح جميع التفاصيل المتعلقة بإعداد البرنامج اللغوي والمحتوى التعليمي الإلكتروني؛ لضمان وضوح الأهداف والمراحل التعليمية. | عدد الوثائق المرجعية المتوفرة | مراجعة الملفات المرجعية |
| | | ب. توافر وثيقة مرجعية لكل المكونات والمقررات الإلكترونية لتعليم اللغة العربية لغة ثانية. | يُستحسَن كتابة وثيقة مرجعية لكل مقرر إلكتروني أو مكوّن من مكونات تعليم اللغة، تتضمن الأهداف والمحتوى والمخرجات المتوقعة؛ مما يسهل على المعلمين والمتعلمين متابعة التقدم. | نسبة المقررات التي لها وثائق مرجعية | مراجعة الأدلة المرجعية للمقررات |
| **ثامنًا: المعايير التقنية** | ١. توفير بنية تحتية داعمة للمنصة | أ. ضمان جودة أداء النظام وسرعة الاستجابة. | التأكد من سرعة استجابة النظام عبر تحليل زمن التحميل ومعالجة الطلبات، وتحسين أداء العمليات المتنوعة في المنصة، مثل: ضغط الملفات، وتوزيع الحمل على الخوادم. | متوسط زمن الاستجابة بالثواني | اختبار أداء الخوادم وأدوات مراقبة السرعة |
| | | ب. التأكد من تحقق التوافقية مع الأجهزة المختلفة. | ضمان توافق تصميم المنصة مع مختلف الشاشات عبر التصميم المتجاوب، ودعم المتصفحات وأنظمة التشغيل المختلفة. | نسبة التوافق مع الأجهزة | اختبارات على مختلف الأجهزة والمتصفحات |
| | | ج. التأكد من قابلية المنصة للتوسع والتحديث. | تصميم المنصة ليكون قابلا للتوسع، بهيكلة نمطية تعتمد على فصل المكونات البرمجية. | عدد التحديثات أو الوحدات المضافة دون تعطيل | مراجعة سجل الصيانة وتحديث المنصة |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | د. ضمان الموثوقية والاستقرار في أنظمة المنصة. | التأكد من استقرار أنظمة المنصة عبر معالجة الأخطاء والاستعادة التلقائية للأنظمة، والتحقق من خلال اختبارات الأداء. | نسبة زمن التشغيل (Uptime) | تقارير المراقبة التقنية |
| | ٢. تحقيق أعلى درجات الأمان والخصوصية | أ. تطبيق أنظمة إجراءات حماية بيانات المستخدمين. | ضمان حماية بيانات المستخدمين عبر تبني تشفير البيانات، وتطبيق سياسات الخصوصية وحماية البيانات الشخصية، ومعالجتها وفق أفضل الممارسات. | عدد سياسات حماية البيانات المفعّلة | مراجعة السياسات وتدقيق الأمان |
| | | ب. تطبيق إجراءات التشفير والتأكد من ضمان أمن المعلومات. | تطبيق بروتوكولات وإجراءات الأمان والحماية من الهجمات وتأمين الواجهات. | نسبة البيانات المشفرة | اختبار بروتوكولات SSL/HTTPS |
| | | ج. تطبيق صلاحيات الوصول والتحكم وفقًا للمعايير العالمية. | تحقيق أفضل الإجراءات في إدارة الهويات، ومستويات الوصول، ومراقبة الأنشطة | عدد مستويات الصلاحيات المطبقة | تدقيق إدارة الهويات والصلاحيات |
| | | د. تطبيق إجراءات النسخ الاحتياطي والاسترجاع للبيانات. | ضمان تطبيق إستراتيجية للنسخ، واستعادة البيانات، وتطبيق معايير إدارة التخزين. | نسبة النسخ الاحتياطي الناجح | مراجعة سجلات النسخ الاحتياطي |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | ٣. تحقيق التكامل التقني | أ. ضمان التوافق مع المعايير العالمية ذات العلاقة بالمعايير التقنية. | تطبيق المعايير التقنية مثل معايير الويب HTML5/ CSS3 ومعايير الأمان ISO 27001 وغيرها من المعايير ذات العلاقة. | نسبة الالتزام بالمعايير العالمية التقنية المعتمدة في تصميم وتنفيذ البرنامج أو المنصة | تدقيق فني لمطابقة المعايير |
| | | ب. قابلية أنظمة المنصة للتشغيل البيني، ومشاركة البيانات. | دعم الواجهات البرمجية مثل: (REST APIs. GraphQL. WebSockets)، وتطبيق سياسات تبادل البيانات والتكامل بين الأنظمة. | عدد عمليات التكامل النشطة | مراجعة تقارير الـ API |
| | | ج. دعم التقنيات المساعدة لجميع فئات مستخدمي المنصة. | دعم تقنيات القراءة، مثل قراءة الشاشة والتكبير والتصغير ووسائل الإدخال مثل التحكم الصوتي وتخصيص العرض. | نسبة الخصائص الداعمة للتقنيات المساعدة | اختبار قارئات الشاشة وأدوات الوصول |
| | | د. تحقيق التكامل مع الأنظمة الأخرى ذات العلاقة بوظائف المنصة. | تكامل المنصة مع الأنظمة الأخرى، مثل: أنظمة التسجيل، والمحاسبة، ومنصات التعلم، وخدمات المصادقة، والدفع، وأدوات التقييم. | عدد الأنظمة المتكاملة فعليًّا | مراجعة تكامل المنصة مع الأنظمة الأخرى |
| | ٤. مراعاة ضوابط بناء الوسائط المتعددة | أ. مراعاة توافر جودة الصوت والصورة في المواد التعليمية. | توفير تسجيلات صوتية نقية ومعالجة احترافياً في استوديوهات متخصصة، مع صور وفيديوهات عالية الدقة تتناسب مع مختلف الأجهزة؛ بما يضمن تجربة تعليمية مثالية تمكن المتعلم من تمييز تفاصيل النطق وفهم تعبيرات الوجه والإيماءات بوضوح. | نسبة المواد المطابقة لمعايير الجودة | تدقيق ملفات الوسائط |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ب. توافر التفاعلية والجاذبية في تصميم مواد الوسائط المتعددة. | تقدم المنصة تجربة تعليمية غامرة بمزج المحادثات التفاعلية الذكية، والألعاب اللغوية المشوقة، وتمارين النطق مع التغذية الراجعة الفورية، والتصميم العصري الجذاب، مدعومًا بنظام مكافآت محفز يشجع على مواصلة التعلم. | نسبة المواد التفاعلية | مراجعة محتوى الوسائط |
| | | ج. تكامل الوسائط المختلفة؛ لتحقيق الإفادة القصوى منها في تحقيق الأهداف التعليمية. | تكامل وتناغم عناصر الوسائط المتعددة (النص والصوت والصورة والفيديو والرسوم المتحركة)، بأن تعمل جميعها معًا بتزامن دقيق وترابط منطقي؛ لتقديم محتوى تعليمي متكامل يخاطب جميع حواس المتعلم، ويعزز الفهم والاستيعاب. | نسبة الأنشطة التي توظف أكثر من وسيط | تحليل تصميم الأنشطة |
| | | د. تحقُّق الفعالية التعليمية للمواد متعددة الوسائط؛ لتحقيق أهداف تعلم اللغة. | تحقق المنصة الأهداف التعليمية عن طريق تصميم تعليمي مدروس يوظف الوسائط المتعددة إستراتيجيًا، مع تقييم مستمر لفعالية كل عنصر، وتخصيص المحتوى حسب احتياجات المتعلم، مدعوماً بأنظمة تحليل متقدمة تقيس مدى التقدم والاستيعاب، وتقدم الدعم المناسب عند الحاجة. | نسبة المواد التي تحقق الأهداف التعليمية | مراجعة المحتوى وتحليل نتائج التعلم |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| تاسعًا: المحتوى اللغوي والثقافي | ١. مراعاة التدرج في تقديم المحتوى اللغوي | أ. تحقيق التدرج في تقديم المفردات اللغوية. | التقديم المنظم والمتسلسل للمفردات، يبدأ بالكلمات الرئيسة الأكثر استخدامًا في الحياة اليومية، ثم يتدرج نحو المفردات الأكثر تخصُّصًا وتعقيدًا. تُقَدَّم المفردات الجديدة في سياقات مألوفة للمتعلم، مع التكرار المدروس الذي يضمن ترسيخها في ذهنه، مراعيًا في ذلك القدرة الاستيعابية للمتعلم في كل مرحلة. | قياس مستوى التدرج في تقديم المفردات | مراجعة المحتوى |
| | | ب. تحقق التدرج في تقديم التراكيب النحوية. | عرض القواعد النحوية، بدءًا من الجمل البسيطة والتراكيب الأساسية، وصولاً إلى التراكيب الأكثر تعقيداً. يضمن هذا التدرج فهم المتعلم للقواعد الأساسية قبل الانتقال إلى المستويات المتقدمة، مع توفير فرص كافية للتطبيق العملي وربط القواعد الجديدة بما سبق تعلمه. | قياس مستوى التدرج في تقديم التراكيب | مراجعة المحتوى |
| | | ج. البناء التراكمي لمهارات اللغة وعناصرها بأن يعتمد كل مستوى على ما سبقه. | يرتكز هذا المعيار على مبدأ البناء التراكمي للمهارات اللغوية؛ إذ تعتمد كل مرحلة على المراحل السابقة. يضمن هذا النهج إتقان المتعلم للمهارات الأساسية قبل الانتقال إلى مستويات أعلى، مع دمج المهارات السابقة في التدريبات الجديدة لتحقيق التكامل اللغوي المنشود. | نسبة المهارات التراكمية | حصر المهارات اللغوية في كل مستوى |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | ٢. مراعاة الدقة اللغوية في المحتوى التعليمي | ا. مراجعة المتخصصين في اللغة للمحتوى مراجعة لغوية دقيقة؛ للتأكد من سلامته نحويًا وأسلوبيًا وإملائيًا. | يتطلب هذا المعيار خضوع المحتوى اللغوي لمراجعة متخصصين في اللغة مراجعة دقيقة؛ للتأكد من سلامة التراكيب النحوية وصحة الأساليب المستخدمة. تضمن هذه المراجعة المهنية مناسبة المستوى اللغوي للفئة المستهدفة والالتزام بالمعايير الأكاديمية في تعليم اللغة. | التحقق من مستوى السلامة اللغوية | مراجعة المحتوى لغويا |
| | | ب. الاعتماد على مصادر ومراجع موثوقة ومعتمدة، مع التحقق المستمر من حداثة المعلومات اللغوية والثقافية. | يضمن هذا المعيار اعتماد المحتوى على مصادر موثوقة ومراجع معتمدة، والتحقق المستمر من أن المعلومات اللغوية والثقافية المضمنة فيها هي معلومات حديثة. يشمل ذلك التحديث الدوري للمحتوى وفق أحدث المعايير والدراسات في مجال تعليم اللغات. | نسبة المراجع الحديثة الموثوقة من مجموع المراجع | حصر المراجع وتحليلها |
| | ٣. تمثيل التنوع اللغوي | أ. تمثيل اللهجات المختلفة للغة الهدف بحسب أهداف البرنامج التعليمي. | تقديم تنوع اللهجات في اللغة المستهدفة تقديماً منظماً ومدروساً، مع توضيح السياقات المناسبة لاستخدام كل لهجة. يركز في اللهجات الأكثر انتشارًا وأهميةً للمتعلمين، وفق أهداف البرنامج، مع ربطها بالسياقات الثقافية والاجتماعية المناسبة. | نسبة اللهجات المتوفرة في البرنامج مقارنة بلهجات الفئات المستهدفة | حصر اللهجات المتوفرة ولهجات الفئات المستهدفة |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ب. تمثيل المستويات اللغوية المختلفة (رسمي، عامي، أكاديمي) بحسب أهداف البرنامج التعليمي. | يقدم هذا المعيار اللغة بمستوياتها المختلفة (الرسمية، والعامية، والأكاديمية) وفق أهداف البرنامج التعليمي. يوضح متى وكيف يستخدم كل مستوى، مع تقديم أمثلة واقعية، وتدريب المتعلمين على التنقل بين هذه المستويات بكفاءة. | نسبة توفر المستويات اللغوية مقارنة بأهداف البرنامج | مقارنة المستويات المتوفرة بالأهداف المطلوبة. |
| | | ج. تمثيل التعبيرات الاصطلاحية والثقافية في المحتوى التعليمي اللغوي. | يركز هذا المعيار على تقديم التعبيرات الاصطلاحية والثقافية في سياقات واقعية، مع شرح معانيها المجازية وارتباطاتها الثقافية. ويساعد المتعلمين على فهم وتوظيف هذه التعبيرات في المواقف المناسبة، مع مراعاة الفروق الثقافية في استخدامها. | توفر التعبيرات الاصطلاحية المثرية | مراجعة المحتوى |
| | | د. تمثيل اللغة للسياقات الاجتماعية المختلفة، ودعم الكفايات الثقافية الاجتماعية لمتعلم اللغة الثانية. | يهدف هذا المعيار إلى تمثيل اللغة لمجموعة متنوعة من المواقف الاجتماعية، مع تدريب المتعلمين على اختيار الأسلوب اللغوي المناسب لكل موقف. ويعكس التنوع في التفاعلات الاجتماعية الواقعية ويقدم نماذج عملية للتواصل الفعال. | توفر الاستخدامات اللغوية السياقية في محتوى البرنامج | تتبع المعاني السياقية في المحتوى التعليمي |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | ٤. تكامل المهارات اللغوية | أ. التقديم المتكامل للمهارات اللغوية (الاستماع، والكلام، والقراءة، والكتابة)، وكذلك للعناصر اللغوية (الأصوات، والمفردات، والتراكيب والقواعد). | يركز هذا المعيار في تقديم المهارات اللغوية (الاستماع، والتحدث، والقراءة، والكتابة) تقديماً متكاملاً ومترابطاً. وتصميم الأنشطة والتدريبات بأن تجمع بين المهارات المختلفة في سياق واحد؛ مما يعزز التعلم الشامل والفعال. | نسبة تغطية البرنامج للمهارات اللغوية والعناصر اللغوية | تحليل المحتوى التعليمي |
| | | ب. تطوير الطلاقة اللغوية بالتدريبات التكاملية والأنشطة التفاعلية. | تطوير قدرة المتعلمين على استخدام اللغة بطلاقة وثقة عن طريق أنشطة تفاعلية تحاكي مواقف الحياة الواقعية. والتركيز في تنمية مهارات التواصل التلقائي والفعال في مختلف السياقات. | نسبة الأنشطة التي تحقق الطلاقة اللغوية مقارنة بالمحتوى التعليمي | مقارنة تحليلية |
| | | ج. تقييم مهارات اللغة وعناصرها بناء على أساليب وأدوات تقييم تكاملية. | اعتماد أساليب تقييم شاملة تقيس جميع جوانب الكفاءة اللغوية، مع تقديم تغذية راجعة مفصلة وبناءة لتوجيه عملية التعلم وتحسين الأداء اللغوي للمتعلمين. | توفر أدوات تقييم متكاملة | تحليل المحتوى |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | ٥. توظيف المحتوى الثقافي بطريقة تعزز الاندماج واحترام الثقافات | أ. تقديم العادات والتقاليد وفقًا للبيئة التي يطبق فيها البرنامج التعليمي. | تقديم المحتوى الثقافي بما يعكس العادات والتقاليد المحلية، مع شرح وافٍ للممارسات الثقافية وربطها بالسياق اللغوي؛ لمساعدة متعلمي اللغة على فهم الثقافة المستهدفة وتقديرها. | اشتمال المحتوى على تعليم العادات والتقاليد. | مراجعة المحتوى التعليمي |
| | | ب. مساعدة المستخدم على التعرف والاندماج في السياق الاجتماعي. | مساعدة المتعلمين على فهم المعايير الاجتماعية المناسبة وتطبيقها، وتطوير مهارات التواصل الفعال في المجتمع المستهدف، مع تقديم إستراتيجيات عملية للتفاعل الاجتماعي الناجح. | توفر أنشطة الغمس اللغوي للمتعلم مع المجتمع | التحقق من الأنشطة التعليمية المصاحبة |
| | | ج. ترسيخ قيمة احترام التنوع الثقافي في المحتوى التعليمي اللغوي. | يعزز هذا المعيار قيم التفاهم والاحترام بين الثقافات المختلفة، مع طرح وجهات نظر متنوعة وتشجيع الانفتاح الثقافي، وتجنُّب الصور النمطية، وتعزيز قيم التسامح والتقبل في سياق تعلم اللغة. | ظهور قيمة التنوع الثقافي | مراجعة المحتوى |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | د. ربط تعلم اللغة بالممارسات الاجتماعية والثقافية في اللغة الهدف. | دمج المحتوى اللغوي في سياقات اجتماعية وثقافية حقيقية تعكس الحياة اليومية للناطقين باللغة الهدف؛ مما يساعد المتعلم على فهم اللغة وممارستها ضمن سياقاتها الطبيعية. ويشمل ذلك: العادات والتقاليد والمنابات الاجتماعية وأنماط التواصل والسلوك المقبولة في المجتمع المستهدف، بما يُمكن المتعلم من استخدام اللغة بفعالية في المواقف الحياتية المختلفة، ويعزز كفاءته اللغوية والثقافية معاً. | نسبة المحتوى المربوط بالممارسات الاجتماعية الثقافية مقارنة بالمحتوى العام | دراسة تحليلية |

## المراجع :


- الباوي، ماجدة إبراهيم وغازي، أحمد باسل. (٢٠١٩). أثر استخدام المنصة التعليمية Google Classroom في تحصيل طلبة قسم الحاسبات. *المجلة الدولية للبحوث في العلوم التربوية*، ٢(٢)، ١٢٦-١٧٠.
- جانييه، روبرت م. (٢٠٠٠). *أصول تكنولوجيا التعليم* (ترجمة محمد الشيخ وآخرون). جامعة الملك سعود.
- حجازي، عبد المعطي. (٢٠٠٩). *هندسة الوسائل التعليمية* (ط.١). دار أسامة.
- دحماني، فاطمة. (٢٠١٩). *استخدامات الطلبة الجامعيين للمنصات التعليمية الإلكترونية* [رسالة ماجستير]. جامعة محمد بوضياف المسيلة.
- الشفيع، عمر الشفيع أحمد. (٢٠١٦). *استخدام الوسائل وتكنولوجيا التعليم للتدريب على طرائق التدريس الحديثة* [رسالة دكتوراه]. جامعة السودان للعلوم والتكنولوجيا.
- شنطاوي، إسلام، وعبد الغني، قمر الزمان، ونوح، محمد. (٢٠١٤). تقويم الوسائل التعليمية في منهاج اللغة العربية. *المجلة الأردنية في العلوم التربوية*، ١٠(١)، ٥٥-٦٤.
- الشوارية، دالية خليل عبد الكريم. (٢٠١٩). درجة *استخدام طلبة الدراسات العليا في الجامعات الأردنية الخاصة للمنصات التعليمية* [رسالة ماجستير]. جامعة الشرق الأوسط.
- عثمانوفيك، ناصح. (٢٠٢٣). أهمية المنصات والمدونات الإلكترونية في تعليم اللغة العربية. *مجلة آفاق للأبحاث السياسية والقانونية*، ٦ (١).

- الغامـدي، سـلطان عبـدالله وفلاتـه، أحمـد محمـد. (٢٠٢٣). معايـير مقترحـة لتصميـم المنصـات التعليميـة الإلكترونيـة. دراسـات عربيـة في التربيـة وعلـم النفـس، (١٤٥)، ١٥٩-١٨٢.
- القاضي، هشـام صالـح. (٢٠١٩). اكتسـاب العربيـة لغـة ثانيـة بواسـطة تطبيقـات الجـوال. *مجلـة التخطيـط والسياسـة اللغويـة*، ٤(٨).
- المركـز التربـوي للغـة العربيـة لـدول الخليـج. (٢٠٢٣). *الإطـار المرجعـي لتعليـم اللغـة العربيـة للناطقـين بلغـات أخـرى*. منشـورات المركـز.
- المركـز الوطـني للتعليـم الإلكـتروني. (٢٠٢٠). *معايـير التعليـم الإلكـتروني للتدريـب بالمملكـة العربيـة السـعودية*. منشـورات المركـز الوطـني للتعليـم الإلكـتروني.
- المجلـس الأوروبي. (٢٠١٦). *الإطـار المرجعـي الأوروبي المشـترك لتعلـم اللغـات* (ترجمـة عبـد الناصـر عثمـان). جامعـة أم القـرى.
- الملاح، تامر المغاوري. (٢٠١٧). *المواطنة الرقمية*. دار السحاب.
- معهـد اللغـة العربيـة بجامعـة أم القـرى. (٢٠١٥). *وثيقـة منهـج تعليـم اللغـة العربيـة للناطقـين بلغـات أخـرى*. جامعـة أم القرى.
- اليـوبي، خالـد محمـد حسـين. (٢٠١٩). فاعليـة المواقـع الإلكترونيـة في تعليـم اللغـة العربيـة للناطقـين بغيرهـا. *مجلـة الأثـر*، (٢٩).

- ACTFL. (2012). ACTFL *proficiency guidelines*. American Council on the Teaching of Foreign Languages.
- ACTFL. (2023). ACTFL *performance descriptors for language learners*. American Council on the Teaching of Foreign Languages.
- Adelsberger, H. H., Collis, B., & Pawlowski, J. M. (Eds.). (2015). *Handbook on Information Technologies for Education and Training* (2nd ed.). Springer.

- Advanced Distributed Learning Initiative. (2024). *SCORM 2004 4th edition technical specifications*. ADL Initiative. https://adlnet.gov/projects/scorm/
- Al Faddaa, H., Al Faddaa, M., & Al-Qassim, R. (2020). International standards for e-learning ESL programs: A comparative study. *Revista Argentina de Clínica Psicológica, 29*(5), 174186-.
- Arkorful, V., & Abaidoo, N. (2014). The role of e-learning: The advantages and disadvantages of its adoption in higher education. *International Journal of Education and Research, 2*(12), 397410-. https://www.ijern.com/journal/2014/December-201434/.pdf
- Atrubi, S. (2008). *E-Learning and Information Services* (1st ed.). Dar Al Arabi.
- Ceci, L. (2024, June 24). *Change in global learning app installs on January 1,* 2024 compared to December 2023. Statista. https://www.statista.com/statistics/1441111/learning-app-install-change-new-year/
- Consortium for School Networking. (2014, March 20). CoSN 2014: *How do big data and digital learning improve education?* EdTech Magazine. https://edtechmagazine.com/k12/article/201403//cosn-2014-how-do-big-data-and-digital-learning-improve-education
- Council of Europe. (2023). *Common European Framework of Reference for Languages: Learning, teaching, assessment* (CEFR). Cambridge University Press.
- Flexible Learning Advisory Group. (2003). *VET learning object repository: Green paper for discussion*.
- Habbal, F. (2022). Standards of building e-learning systems in higher education. *Emirati Journal of Business, Economics, and Social Studies, 1*(1), 430-.
- IEEE LTSC. (2024). *IEEE standard for learning technology systems architecture*. Institute of Electrical and Electronics Engineers.
- IMS Global Learning Consortium. (2024). *IMS Question & Test Interoperability (QTI) specification* (Version 3.0). IMS Global.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً


- International Organization for Standardization. (2023). ISO/*IEC 40180:2023 Information technology - Quality for learning, education and training - Fundamentals and reference framework.* ISO.
- International Society for Technology in Education. (2017). *ISTE Standards for Students: A practical guide for learning with technology* (1st ed.).
- Means, B., & Olson, K. (1995). *Technology›s role in education reform: Findings from a national study of innovative schools*. Office of Educational Research and Improvement.
- Müazín, A. D., & Jesry, M. M. (2019). E-Learning platforms and their role in teaching Arabic as a second language: "Kahoot and Global Arabic Open Platform Model". *Route Education and Social Science Journal, 6*(10).
- OECD. (2005). *E-learning in tertiary education*. Organisation for Economic Co-operation and Development.
- Quality Matters. (2023). *Higher education course design rubric* (7th ed.).
- Swender, E. (Ed.). (2012). *ACTFL Certified Proficiency Interview Tester Training Manual*. ACTFL.
- TESOL International Association. (2022). *Standards for technology in language teaching*. TESOL Press.
- TESOL International Association. (2023). *Technology standards framework for language teaching*. TESOL Press.
- World Wide Web Consortium. (2023). *Web Content Accessibility Guidelines (WCAG) 2.2*. W3C.
- World Wide Web Consortium. (2024). *Making content usable for people with cognitive and learning disabilities*. W3C.

# الفصل الثالث

## معايير بناء المعاجم الرقمية العامة

- ملخص
- مقدمة
- مدخل تعريفي
- الدراسات السابقة
- نطاق المعايير وحدودها
- قائمة المعايير
- المراجع



هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

## ملخص

يقدم هذا الفصل مجموعة من المعايير التي ينبغي مراعاتها في بناء المعاجم الرقمية العامة؛ لتسهم في تحسين مستوى جودة المعاجم الرقمية العربية، وتساعد مطوري المعاجم العربية والمهتمين بها على استثمار أحدث التقنيات في كل تفاصيل الصناعة المعجمية. ويبدأ الفصل بمقدمة ثم مدخل تعريفيّ لموضوع المعايير وأهميتها، ثم يعرض مراجعة للدراسات السابقة ذات الصلة بمعايير بناء المعاجم الرقمية وأطرها وأدلتها المرجعية، مع الإشارة إلى أهم تحديات صناعة المعجم الرقمي العربي. ويلي ذلك قسم خاص يعرف بحدود الوثيقة ومجالات تطبيقها، ويختتم الفصل بسرد لقائمة المعايير، مصنفة في فئات متعددة، وفقًا لمراحل بناء المعاجم الرقمية وتصميمها.

## مقدمة

يُعد الحديث حول أهمية المعايير ودورها المحوري في رفع جودة بناء المعاجم بمختلف أنواعها من أكثر الموضوعات شُيوعًا في أدبيات المعاجم عموماً؛ فمراعاة مثل هذه المعايير يضمن تنظيم كل مراحل وعمليات صناعة المعجم وجودتها، ابتداءً من فكرة بناء المعجم حتى الوصول إلى مراحل النشر والتوزيع والمراجعة والتقويم. وتزداد أهمية مراعاة المعايير في بناء المعاجم الرقمية؛ إذ تتضمن صناعتها تفاصيل أكثر وتقنيات أوسع مقارنة مع المتطلبات التقليدية للمعاجم المطبوعة. واتباع المعايير المنضبطة يضمن جودة جميع مكونات صناعة المعاجم الرقمية، مثل: الكتابة البرمجية، والتحرير المعجمي، والتوسيم اللغوي، مع تحقيق الفعالية والكفاءة والكفاية، والإفادة القصوى من السمات والأدوات التقنية المتنوعة، في جميع تفاصيل العمل المعجمي الرقمي.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

وتتجلى أهمية توظيف المعايير والوثائق المرجعية في إنتاج وصناعة المعاجم الرقمية من خلال النقاط الآتية:

- **ضمان الاتساق والوضوح في بناء المعجم:** يساعد تحقيقهما -في جميع عمليات صناعة المعجم- على الوصول إلى نظام معجمي دقيق، يلبي الأغراض المرجوّة، ويسهم في توحيد إجراءات الصناعة المعجمية وضبطها، كما يتيح سهولة الوصول للمعلومات ونشرها وإعادة استعمالها بطرق متعددة.

- **تيسير إعادة الاستعمال وتبادل البيانات المعجمية:** يمكن دمج المعاجم الإلكترونية المبنية على المعايير العالمية الثابتة -بسهولة- في مختلف الأنظمة الحاسوبية، وكذلك إعادة استعمالها لأغراض متعددة، ونشرها إلكترونيًّا في قوالب متعددة بأقل جهد وتكلفة ممكنة.

- **التعاون والفهم المشترك:** تساعد المعايير على توحيد الجهود وتشجيع التعاون بين المختصين والجهات ذات العلاقة في بناء المعاجم الإلكترونية، لما تقدمه من قوالب ونماذج حاسوبية مشتركة، وتقنيات تعاونية، تسهم في إنجاز الأعمال المعجمية الجماعية بطرق وآليات مختلفة، ولا سيما في ميدان المعاجم المختصة.

- **تحقيق الفعالية والكفاءة وزيادة الإنتاجية:** وذلك بتقليل احتمالية تكرار العمل، وتشتت الأهداف وتداخلها، وتضاعف التكلفة والوقت اللازم لإصلاحها.

- **رفع مستوى الجودة والموثوقية:** إذ يحقق المعجم المبني على المعايير ضماناً للجودة؛ مما يجعل الجهات والأفراد يعتمدون عليه، ويثقون به.

- **فتح الباب لدراسات وبحوث المعاجم:** يمكن استعمال بيانات هذه المعاجم في تطبيق عدد من الدراسات المعجمية لأغراض تطوير وابتكار تقنيات ومنصات إدارة المعاجم والبحث فيها؛ إذ يسهل الاستشهاد بها وتحليل بياناتها، ومقارنتها مع عدد من المعاجم المشابهة.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

**مساعدة الجهات الإشرافية على ضمان تطبيق المعايير ومراعاتها:** يخضع العمل المعجمي في كثير من الجهات المحلية والدولية لمتطلبات خاصة تضمن تحقيق أهدافها، ومن شأن هذه المعايير أن تساعد تلك الجهات، باستعمالها في مراحل قبول وتحرير ونشر المعاجم الإلكترونية، كما يمكن اعتمادها لتكون مرجعًا رئيسًا للأنظمة والسياسات الرسمية التي قد تعد مستقبلًا للتعامل مع البيانات المعجمية

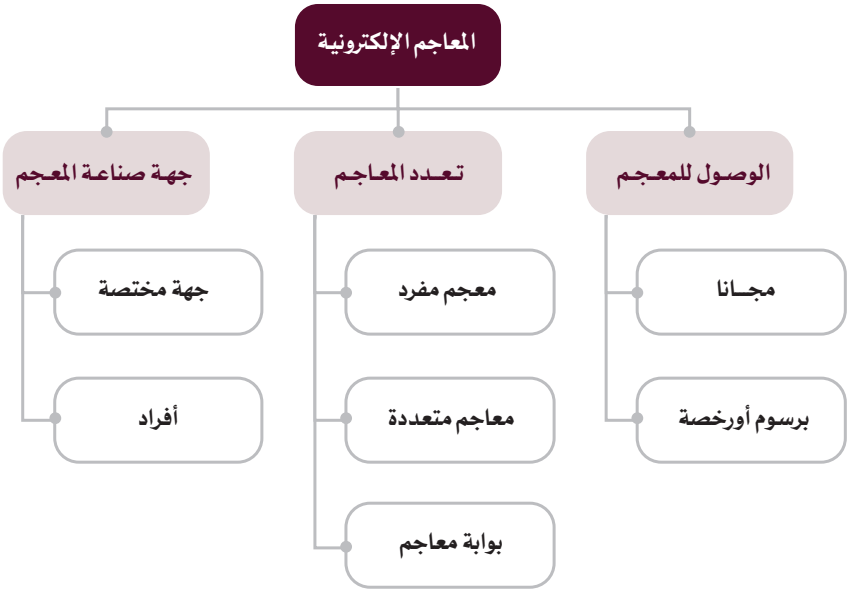
**لأجل هذه الأسباب وغيرها:** فقد حظيت الأطر الإرشادية المرجعية ووثائق المعايير المحلية والعالمية باهتمام متزايد من المؤسسات المعنية بضبط جودة الإنتاج المعجمي، وتعزيز الصناعة المعجمية عموماً؛ إذ تُيَسِّر مشاركة المعلومات المعجمية وتُسهِّل التعاون في جمعها وصناعتها على المستويين المحلي والدولي. وقد أدرك كثير من المختصين مدى تعقيد صناعة المعاجم وتعدد التخصصات الداخلة فيها، ومن ثَمّ ضرورة بذل الجهود في اقتراح أطر ومعايير تساعد على مشاركة البيانات المعجمية والإفادة منها في تطبيقات متعددة (Müller-Spitzer, 2008; Romary & Wegstein, 2012; Pilehvar & Navigli, 2014; McCracken, 2016; McCrae et al., 2019; Salgado et al., 2019, among others).

**وفي عصر الاتصال والمعلومات** لم يعد التحدي في توافر البيانات عند الحاجة إليها، بل في الحصول عليها في القالب والشكل المناسب للمستخدم؛ فكمية البيانات الضخمة المتوافرة قد تكون سببا في صعوبة الوصول إلى المعلومة المناسبة في الوقت المحدد. وذلك يفرض على المعجميين استثمار الأدوات التقنية المتاحة لتحقيق وصول سريع وكافٍ ومباشر للمعلومات المعجمية الموثوقة والمناسبة والمحدثة باستمرار، بدلًا من البحث في أكوام البيانات الضخمة المتاحة عبر الإنترنت، التي قد تضيع الجهد والوقت دون الوصول إلى التعريف والمعنى المناسب للكلمة أو التركيب. ومن هنا يقدم هذ الفصل مجموعة من المعايير الرئيسة التي يلزم مراعاتها عند صناعة المعاجم الإلكترونية لأغراض عامة وإعدادها.

## مدخل تعريفي:

عـرَّف جرانجـر وباكـوت (Granger & Paquot, 2012) المعجـم الإلكـتروني بأنـه: عمليـة جمع البيانـات المعجميـة مـن مصـادر ومـوارد لغويـة إلكترونيـة، وتنظيمهـا في مداخـل معجميـة لأغـراض متعـددة، مـع ترميزهـا حاسـوبيًّا، وإضافـة معلومـات إضافيـة ذات علاقـة، مثـل: المعـاني، وأمثلـة الاسـتعمال، والترجمـة، مـع إمكانيـة تخزيـن بياناتـه محليًّـا أو عـبر الإنترنـت، ونشـره في قوالـب إلكترونيـة متنوعـة، تتضمـن الحاسـوب بـكل أنواعـه والهواتـف الذكيـة والأجهـزة اللوحيـة وغيرهـا.

وصنفـت بعـض الدراسـات (De Schryver, 2003; Nesi, 2009; Lew, 2011a) المعجـم الإلكـتروني إلى فئـات متعـددة، كمـا هـو موضـح في (الشـكل١)، حيـث يمكـن تصنيـف المعاجـم أو القواميـس وفقًـا لاعتبـارات متعـددة، منهـا: إمكانيـة الوصـول للمعاجـم، ومسـتوى تعـدد المعاجـم، والجهـات القائمـة علـى بنـاء المعاجـم وتطويرهـا.




الشكل ١: أمثلة لتصنيفات وأنواع المعاجم الإلكترونية.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

وأشـار سـالجادو (Salgado, 2021) إلى أن المعجـم الإلكـتروني يمكـن النظـر إليـه مـن زوايـا متعـددة؛ فهـو قـد يُعـدّ نصًّـا لغويًّـا، أو مشـكلة بحثيـة، أو أثـرًا ثقافيًّـا، أو أداة تعليميـة، أو نموذجًـا لغويًّـا حاسـوبيًّا. وقـد بـدأ الحديـث عـن أهميـة المعاييـر في ضبـط الجـودة وضمانهـا في صناعـة المعاجـم الإلكترونيـة منـذ السـبعينيات والثمانينيـات مـع بدايـة انتشـار اسـتعمال المعاجـم الإلكترونيـة عـن طريـق الحواسـيب الشـخصية؛ فقـد ظهـرت حينئـذٍ محـاولات مبكـرة لإعـداد معاييـر عامـة لبنـاء المعاجـم الإلكترونيـة. وفي الوقـت الحاضـر تتجـه معظـم المعاجـم إلى النشـر الإلكـتروني حـتى لـو صُممـت ابتـداءً للنشـر الـورقي؛ للفوائـد الجمَّـة والسـمات المميـزة للمعاجـم الإلكترونيـة، الـتي مـن أبرزهـا: تقنيـات البحـث الإلكترونيـة، وسـهولة الوصـول، وإمكانـات التخزيـن المحلـي والسـحابي للبيانـات المعجميـة.

ويقصـد في سـياق هـذه الوثيقـة بـ "المعجـم العـام": المعجـم الإلكـتروني العـام الـذي يقابـل المعجـم المختـص أو الاصطـلاحي، ويكـون محتـواه اللغـوي عامًّـا ومُمثـلًا للاسـتعمال اللغـوي العـام، دون اختصاصـه بمصطلحـات حقـل معـرفي أو مجـال مهـني معـين. وتتضمـن المعاجـم العامـة جميـع أنـواع المداخـل المعجميـة، مثـل: المفـردات، والمصطلحـات العامـة، والتعبـيرات والألفـاظ الوظيفيـة. وتشـمل الوثيقـة جميـع المعاييـر الواجـب مراعاتهـا في بنـاء المعاجـم الإلكترونيـة الموجهـة للاسـتعمال البشـري أو الآلي. ولا يشـمل نطـاق هـذه الوثيقـة: المعاجـم المطبوعـة، والمعاجـم المعـدَّة لأغـراض بحثيـة خاصـة محـدودة التـداول، وقوائـم المفـردات العامـة أو الاصطلاحيـة المُجَمَّعـة لأغـراض خاصـة. كمـا أنهـا لا تشـمل مـا يتعلـق بالموسـوعات المعرفيـة والثقافيـة، أو بنـاء المدونـات، أو قواعـد البيانـات اللغويـة.

ويشـير (الجـدول ١) إلى بعـض الفـروق بـين المدونـات أو قواعـد البيانـات اللغويـة أو المـوارد اللغويـة العامـة، والمعاجـم الإلكترونيـة (Bergenholtz, 2013). ويمكـن القـول إن المدونـات والمـوارد اللغويـة هـي الإطـار الأوسـع والأشـمل للمـوارد اللغويـة، وتمثـل المـادة الخـام الـتي تسـتخرج منهـا المـادة اللغويـة المناسـبة لأهـداف المعجـم الإلكـتروني.

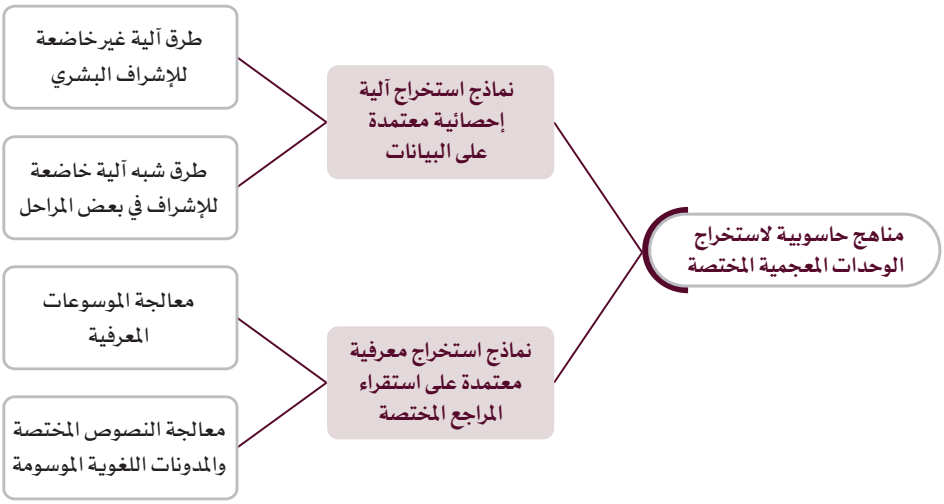
هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| م | الموارد اللغوية / وجه الفرق | قاعدة البيانات اللغوية أو المدونة | المعجم الإلكتروني |
|---|---|---|---|
| ١ | غرضها الرئيس | جمع وتوثيق أكبر قدر من البيانات اللغوية في أنماط وقوالب مختلفة. | أداة تعليمية موجهة لتلبية احتياجات مستخدم محدد، وتعتمد في مصادرها على جمع بيانات محددة من المدونات أو قواعد البيانات اللغوية. |
| ٢ | حجم المعلومات | تتميز بضخامة حجم المعلومات وتوسعها. | تتميز بتركيز المعلومات ودقتها. |
| ٣ | جودة المعلومات | تتفاوت فيها مستويات جودة المعلومات وطرق تنظيمها. | تتميز بمعيارية جودة المعلومات واتساق طرق تنظيمها. |

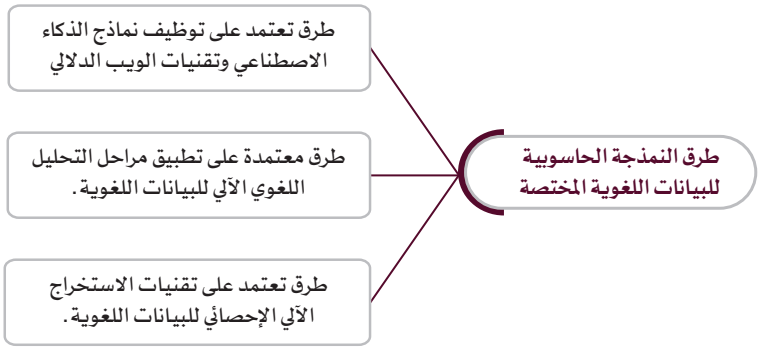
الجدول ١: أمثلة لأهم الفروق بين قواعد البيانات اللغوية والمعاجم الإلكترونية.

وقد تعددت المناهج المعتمدة لاستخراج المصطلحات أو الوحدات المعجمية المختصة من البيانات اللغوية الضخمة. ويوضح (الشكلان ٢و٣) بعض المناهج المستعملة في استخراج الوحدات المعجمية المختصة، فبعضها يعتمد على الإحصاء ومعالجة البيانات اللغوية، وبعضها يعتمد على المعلومات المعرفية والمعاجم المختصة والمدونات اللغوية الموسومة. كما تختلف طرق النمذجة الحاسوبية لهذه البيانات، بين طرق معتمدة على خوارزميات الذكاء الاصطناعي والنماذج اللغوية الضخمة، وطرق معتمدة على المعالجة والتحليل اللغوي الآلي، وأخرى معتمدة على تحليل البيانات اللغوية إحصائيًا، ومحاولة الإفادة من كل ذلك في التعرف الآلي على المصطلحات.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً




الشكل ٢: أهم النماذج الحاسوبية المستعملة في استخراج الوحدات والمصطلحات المعجمية المختصة.




الشكل ٣: أمثلة لطرق النمذجة الحاسوبية للبيانات اللغوية المختصة.

ويمكن تصنيف المعاجم المختصة وفقًا لاعتبارات وعوامل متعددة، **من أهمها:**

- أهـداف المعجـم: فبعـض المعاجـم المختصـة معـدَّة لغـرض التعليـم والشـرح، في حـين تقـدم بعـض المعاجـم الأخـرى للمتخصصـين.
- فئات مستعملي المعجم: ويمكن أن تعد المعاجم المختصة لفئات متعددة، مثل: العامة، أو المتخصصين، أو المترجمين، أو الكتاب واللغويين.
- طريقـة تنظيـم المداخـل المعجميـة: يمكـن أن يعتمـد المعجـم المختـص تقديـم المصطلـح ثـم المفهـوم أو التعريـف الخـاص بـه، أو يعتمـد على تقديـم المفاهيـم أولًا مـع ربطهـا بالألفـاظ الـتي تسـتعمل في الدلالـة عليهـا.
- طبيعـة البيانـات المعجميـة: تميـل بعـض المعاجـم إلى الدقـة والاختصـار في الدلالـة على معـاني المصطلحـات، في حـين تفضـل بعـض المعاجـم المختصة التوسـع في شـرح المفاهيـم وربطهـا بمزيـد مـن المعلومـات الإضافيـة لأغـراض متنوعـة.

وتهـدف أغلـب المعاييـر والوثائـق المرجعيـة الـتي صممـت لتمثيـل البيانـات المعجميـة إلى تحقيـق مبـادئ رئيسـة أشـار إليهـا (Wilkinson et al. ٢٠١٦)، وهـي: سـهولة الوصـول إلى تلـك البيانـات المعجميـة، وإمكانيـة البحـث فيهـا، وقابليتهـا للمعالجـة الحاسـوبية التكامليـة مـع المـوارد اللغويـة المشـابهة، وإعـادة اسـتعمالها لأغـراض مختلفـة، وسـهولة مشـاركتها في التطبيقـات اللغويـة الحاسـوبية المختلفـة.

## الدراسات السابقة

يقـدم هـذا القسـم التطـور التاريـخي لتأليـف وإعـداد الأطـر العامـة والوثائـق الإرشـادية والمعاييـر الخاصـة بالمعاجـم الإلكترونيـة، مـع تحليـل عـدد مـن الوثائـق العالميـة المعتمـدة في هـذا المجـال، ويقـارن واقـع التأليـف والمشـاريع المُعـدة خصوصـاً للغـة العربيـة مـع الوثائـق والأطـر العالميـة. وسـنبدأ بعـرض خلفيـة تاريخيـة للوثائـق والمعاييـر السـابقة، يتبعهـا عـرض موجـز لأهـم الوثائـق المتاحـة.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

فقد بدأ الاهتمام في ثمانينيات القرن العشرين بإعداد معايير ووثائق مرجعية تنظم عمليات بناء المعاجم الإلكترونية العامة، سواء لأغراض الاستعمال البشري أم الآلي. وصاحب هذا الاهتمام عدد من مشاريع المعاجم الإلكترونية متعددة اللغات مثل: (MULTILEX– GENELEX – ACQUILEX). وقد تصدرت هذه المرحلة جهود لجنتين متخصصتين في إعداد معايير لبناء المعاجم الإلكترونية، **هما:**

- مجموعة المستشارين الخبراء لهندسة اللغة (Expert Advisory Group for Language Engineering – EAGLES).
- مجموعة المعايير الدولية لهندسة اللغة (International Standards for Language Engineering – ISLE).

واقترحت هاتان المجموعتان مبادئ منهجية لإعداد المعايير العامة لتصميم وإعداد المعاجم الرقمية؛ بهدف تنسيق الجهود الدولية وتوحيدها في هذا الإطار. ومن أبرز نتائج المرحلة ما قدمته مجموعة (ISLE) من نموذج أولي لتمثيل المداخل المعجمية متعددة اللغات لأغراض عامة (Multilingual ISLE Lexical Entry – MILE)، وقد استند إلى عدد من الدراسات الاستقصائية طويلة المدى هدفت إلى استقراء الطرق المشتركة بين اللغات في تمثيل البيانات المعجمية، وقد نتج عن هذه الدراسات اقتراح مجموعة من المعايير الموحدة لنموذج بناء وتمثيل بيانات المعجم الإلكتروني.

وفي الوقت نفسه بدأت المنظمات الدولية للمقاييس والمعايير–مثل: (ISO – IEC –ITU)– بإعداد وثائق مشابهة وأكثر رسمية في معايير بناء المعاجم الإلكترونية وتمثيل بياناتها (Šimunić & Pavić, 2020). وكانت منظمة الآيزو من أكثر المنظمات اهتمامًا بتوحيد الجهود في هذا المجال؛ إذ بنت على كل الجهود السابقة في إعداد معايير المعاجم الإلكترونية، وتوجتها بإصدار إطار مرجعي موسع لتمثيل البيانات المعجمية فيما يعرف بـ "الإطار المرجعي لتوسيم البيانات المعجمية" (Lexical Markup Framework – LMF). وكانت هذه الجهود المتواصلة في إعداد وثائق المعايير الخاصة بالمعاجم الإلكترونية مدفوعة بالتطور في تقنيات الويب الدّلالي (Semantic Web Technologies)، وتزايد الحاجة إلى المعاجم

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً



الإلكترونية متعددة القوالب في تطبيقات المعالجة اللغوية الحاسوبية متعددة الأغراض. إضافة إلى مواجهة تحديات التعدد اللغوي للمصادر المعجمية، الذي عززه سهولة التواصل وتبادل البيانات عبر الإنترنت. وتُعدّ دراسة خاتشادوريان (Khatchadourian et al., 1994) محاولة مبكرة في تأكيد أهمية إعداد المعايير في جودة بناء المعاجم الإلكترونية وتسهيل إعادة استعمالها لأغراض متعددة.

ويمكن تصنيف تلك المعايير الخاصة بالمعاجم إلى فئتين: معايير تركز في الاستعمال الحاسوبي في عرض المداخل المعجمية، ومعايير أخرى تضع في أولوياتها المستخدم البشري من حيث عرض المداخل والبيانات المعجمية وتنظيمها. ومن أشهر الوثائق التي تمثلها وثيقة معايير الآيزو رقم ١٩٥١ المنشورة (٢٠٠٧) وحدثت عام (٢٠٢١). كما يمكن تصنيف وثائق المعايير بحسب النموذج الرئيس للمعجم إلى معاجم الألفاظ ومعاجم المعاني، وعلى الرغم من رسوخ استعمال هذا التقسيم في التراث المعجمي العربي فإنه لم يناقش في الأدبيات المعجمية الغربية إلا في أوائل القرن العشرين.

ولقد تمثل الدافع المشترك لوثائق المعايير المعجمية في محاولة الوصول إلى إجماع حول تمثيل المحتوى اللغوي وغيره للمعجم، إضافة إلى توحيد طريقة العرض وترميز البيانات المعجمية. **كما تشترك وثائق معايير المعاجم في تحقيق الأهداف الإضافية التالية:**

- توحيد تصميم نموذج البنية التحتية للمعاجم الإلكترونية.
- مساعدة النموذج على تصميم وبناء وتقويم المعجم وفقا لاحتياجات المستخدم النهائي.
- تمكين إعادة الاستعمال، وسهولة الدمج والنقل للبيانات المعجمية.
- تحسين إمكانية سهولة الوصول للمعاجم الإلكترونية وتطويرها المستمر.
- تيسير عمل تطبيقات المعالجة الآلية للغة، ولا سيما تطبيقات الاستخراج الآلي للمعلومات والمصادر المعجمية.
- المساعدة على تقويم المعاجم الإلكترونية، ورفع مستوى تمثيلها وجودتها.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

ومـن نمـاذج المعايـير المبكـرة: النمـوذج العـام للمعاجـم القابلـة لإعـادة الاسـتعمال، الـذي قدمـه أنتـوني لاي وآخـرون (Antoni–Lay et al. 1994)، وصُمِّـم للمعاجـم واسـعة الانتشـار، بهـدف تيسـير إعـادة اسـتعمالها وبنائهـا وتطويرهـا لأغـراض حاسـوبية مختلفـة. وكان النمـوذج جـزءًا مـن مشـروع أكـبر لتطويـر منصـات وأدوات رقميـة لإدارة وتأليـف المعاجـم الإلكترونيـة، ويمكـن عـن طريـق هـذا النمـوذج تمثيـل أنـواع متعـددة مـن البيانـات اللغويـة والعلاقـات التركيبيـة وفقًـا للنظريـات المعجميـة المختلفـة.

واعتمـد مطـورو النمـوذج علـى بيانـات معجميـة تمثـل عـددًا مـن اللغـات الأوروبيـة، مثـل: الإيطاليـة والإسـبانية والبرتغاليـة والفرنسـية. وكان الدافـع الرئيـس تيسـير نقـل البيانـات والمصـادر المعجميـة بـين هـذه اللغـات، مـع تطويـر أدوات حاسـوبية لإدارة ومعالجـة تلـك البيانـات بمـا يحقـق تطـور تطبيقـات المعالجـة الآليـة للغـات البشـرية، إضافـة إلى إثـراء المعاجـم الإلكترونيـة بالمعلومـات اللغويـة. ويُمَثِّـل النمـوذج النظـري البيانـات اللغويـة للمعجـم عـبر ثلاثـة مسـتويات: المسـتوى الصـرفي ثـم النحـوي ثـم الـدلالي، إضافـة إلى تمثيـل العلاقـات المتداخلـة بـين هـذه المسـتويات.

وفي عـام ١٩٩٩ قـدم سـانفيليبو وآخـرون (Sanfilippo et al., 1999) دليـل ترميـز البيانـات المعجميـة والدلاليـة، الـذي شـكّل إطـارًا مرجعيًّـا لتمثيـل البيانـات الدلاليـة في المعاجـم الإلكترونيـة وتطبيقـات المعالجـة الآليـة للغـة، واعتمـد الدليـل علـى نتـائج اجتماعـات مكثفـة شـارك فيهـا أكـثر مـن ٢٥ خبـيرًا. وتضمـن عـددًا مـن الوثائـق المرجعيـة المهمـة لتطبيقـات الترجمـة الآليـة ومـا يشـبهها مـن أنظمـة معالجـة المعلومـات، إضافـة إلى معايـير خاصـة بتحديـد وتنظيـم العلاقـات الدلاليـة للمداخـل المعجميـة، وترتيـب وتصميـم الموسـوعات المعرفيـة الحاسـوبية (الأنطولوجيـا) وتصنيفهـا. إضافـة إلى المعايـير الخاصـة بالبيانـات اللغويـة، مثـل: السـمات الصرفيـة والتركيبيـة المتنوعـة للمدخـل المعجـمي.

وصنـف الدليـل المعلومـات المعجميـة الدلاليـة إلى ثـلاث فئـات وفقًـا لمسـتوى أهميتهـا في تطبيقـات التقنيـات اللغويـة، مـع اهتمـام خـاص بذكـر المعلومـات المعجميـة المصاحبـة

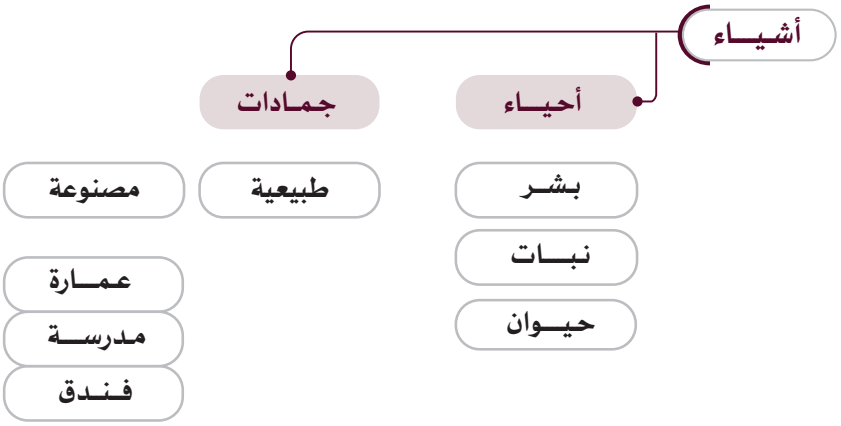
للمدخل المعجمي، مصنفة إلى متوسطة وعالية الأهمية، كما يظهر في (الجدول ٢) الذي يُظهر أمثلة للمعلومات المعجمية في كلا الفئتين.

| البيانات الدلالية | درجة أهميتها |
|---|---|
| المعنى الدلالي الرئيس | عالية |
| علاقة الترادف والتصنيف المفاهيمي | |
| الإطار الدلالي | |
| علاقة الجزئية | متوسطة |
| علاقة التضاد | أمثلة التطبيقات الحاسوبية اللغوية: |
| المجال الموضوعي | — الترجمة الآلية. |
| التصاحب اللفظي الإحصائي | — الاستخراج الآلي للبيانات. |
| الصيغة الفعلية | — التلخيص الآلي للنصوص. |
| العلاقة الزمانية | — التصنيف الآلي للكلمات. |
| جانب زمن الفعل | — الاستخراج الآلي للعبارات الاصطلاحية. |
| الكمية | — التمييز الدلالي للمفردات. |

الجدول ٢: البيانات الدلالية (متوسطة وعالية الأهمية) المصاحبة للمداخل المعجمية

وعلى الرغم من عدم اشتمال هذا الدليل على تطبيقات خاصة باللغة العربية؛ فإنه يمكن الإفادة من بعض المعايير المذكورة فيه، ولا سيما المعايير المتعلقة بالتصنيف الهرمي للعلاقات المفاهيمية، كما في المثال المعروض في (الشكل ٤) الذي يتضمن بعض الأمثلة للعلاقات بين المفاهيم. وقدم الدليل قوائم متعددة من المعايير لكيفية تمثيل البيانات المعجمية الدلالية بأنواعها المختلفة، واقترح عددًا من الضوابط في تنظيمها ونقلها من لغة إلى أخرى.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً




الشكل ٤: مثال للعلاقات المفاهيمية في دليل ترميز البيانات المعجمية والدلالية.

وفي عـام ٢٠٠٠ قـدم لينـسي وآخـرون (Lenci et al., 2000) إطـارًا نظريًّـا عامًّـا لتطويـر المعاجـم متعـددة اللغـات؛ لتلبيـة احتياجـات تطبيقـات المعالجـة الآليـة للغـات الخاصـة ببنـاء المعاجـم الحاسـوبية الدلاليـة، وقـد تزامـن هـذا الجهـد مـع تنـامي الاهتمـام بالنمـاذج الحاسـوبية الخاصـة بفهـم اللغـة الطبيعيـة وإنتاجهـا، وعـدم الاكتفـاء بنتـائج المعالجـة الإحصائيـة لنمـاذج معالجـة المهمـات اللغويـة. وارتبـط الإطـار بمشـروع معجـمي مصاحـب شـمل أكـثر مـن ١٢ لغـة، **وهـدف الإطـار المرجعـي إلى تحقيـق ٣ أهـداف رئيسـة:**

- شرح الدوافع الرئيسة للإطار المرجعي الخاص ببناء المعاجم الحاسوبية.
- توضيح الأدوات والمكونات الدلالية المُستعملة في تمثيل معاني المفردات في المعجم.
- توضيـح آليـة عمـل النمـاذج المعجميـة ووظائفهـا المتنوعـة في مرحلـة تطويـر وبرمجـة المعاجم.

وقدمـت الوثيقـة معايـير عديـدة يمكـن تبنيهـا في الوثيقـة الحاليـة، مثـل: معايـير اختيـار المعـاني، والحقـول الدلاليـة، وطـرق تنظيـم العلاقـات المختلفـة داخـل المعجـم الإلكـتروني. وفيمـا يتعلـق بتحديـد المفاهيـم العامـة للمعجـم فقـد تبنّـت الوثيقـة -مـع بعـض التعديلات-

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً



المفاهيم الخاصة بشبكة المفردات الأوروبية (EuroWordNet). ويهدف النموذج الخاص بتمثيل المعلومات الدلالية الخاصة بمعاني المفردات إلى الإجابة عن **الأسئلة التالية:**

- ما المعلومات الدلالية اللازم ذكرها؟
- كيف تُمثَّل في المعجم الإلكتروني؟
- ما مقدار التفاصيل الدلالية المناسبة للترميز في المعجم المحوسب؟

وتضمن نموذج المعجم الخاص بالمعلومات الدلالية مجموعة من البيانات، على سبيل المثال: الحقل الدلالي، ونوع اللغة، والتراكيب ذات العلاقة، ودرجة شفافية المعنى أو المعاني الاصطلاحية والمجازية المرتبطة به. كما تضمنت الوثيقة ترميز أنواع من المعلومات اللغوية للوحدات المعجمية، مثل: نوع الوحدة المعجمية من حيث الإفراد أو التركيب، إضافة إلى المعلومات الخاصة بالمشتقات الصرفية، وأصل الكلمة، وتحليلها الصرفي.

وأضافت الوثيقة ترميزًا خاصًّا بأنطولوجيا تابعة للمعجم الإلكتروني، تتضمن تصنيف المفردات وفقًا لمجموعة من المفاهيم والعلاقات الدلالية المختلفة. وصُنفت البيانات الملحقة بالمداخل المعجمية إلى ٣ فئات **كما يلي:**

- البيانات الضرورية (ينبغي إضافتها للمدخل المعجمي).
- البيانات المفضلة المُوصى بإضافتها (يفضل إضافتها إلى المدخل المعجمي).
- البيانات الاختيارية (يختارها مطور المعجم وفقًا للسياق والاحتياجات المتنوعة).

وتمثل البيانات الضرورية، الحد الأدنى من البيانات لكل معنى دلالي، يمكن لصانعي المعجم -عن طريقها- التمييز بين المعاني الدلالية المختلفة للوحدة المعجمية. وكل نوع من هذه البيانات يمثل مستوى في النموذج التمثيلي للمداخل المعجمية. ويوضح (الجدول ٣) أهم أنواع البيانات الخاصة بالمستويات الثلاثة.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المستوى | نوع البيانات | المستوى | نوع البيانات |
|---|---|---|---|
| البيانات الضرورية | نوع النموذج<br>Template Type | البيانات المفضلة | النوع الهرمي للبيانات<br>Type hierarchy information |
| | المجال المعرفي للنطاق<br>Domain information | | التراكيب المتصلة<br>Qualia Structure |
| | الصنف الدلالي<br>Semantic class | | المشترك اللفظي<br>Polysemous class |
| | توضيح<br>Glossa | | الجذر والمشتقات الصرفية<br>Derivation |
| | تمثيل تركيبي<br>Predicative Representation | | نوع الحدث<br>Event type |
| | مقيدات مختارة<br>Selectional Restrictions | البيانات الاختيارية | المتصاحبات اللفظية<br>Collocations from the corpus |
| | نوع الصلة بشبكة المعاني الدلالية<br>Link to the SynU | | المرادفات<br>Synonymy |

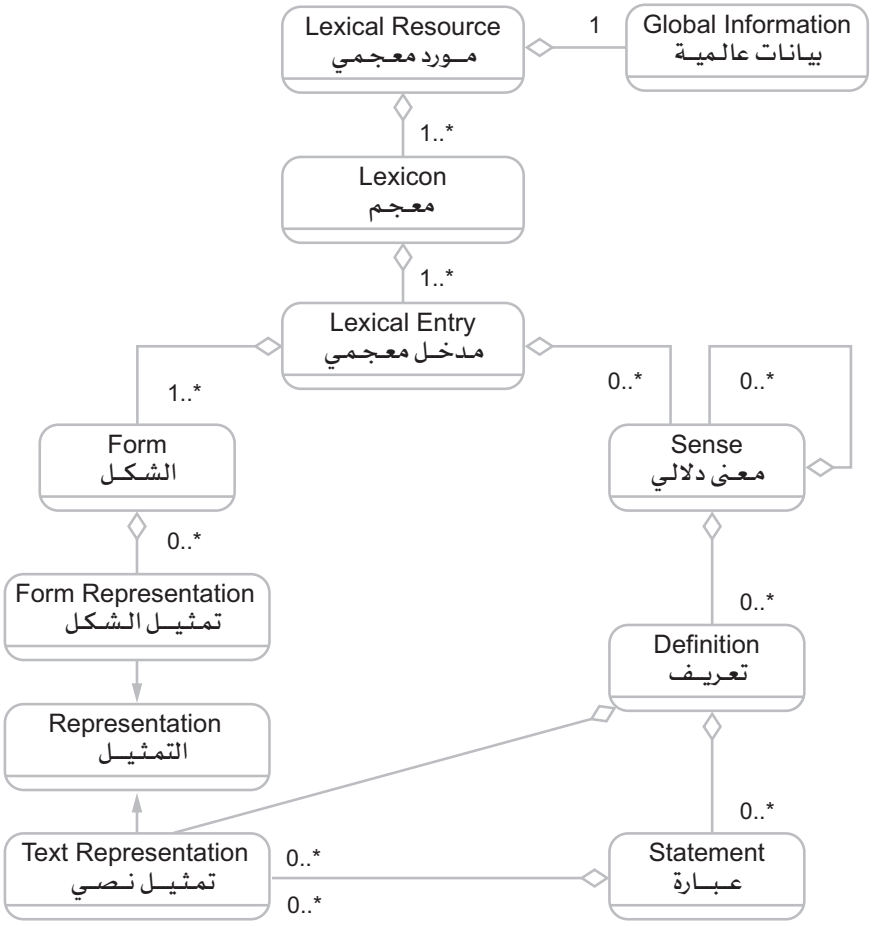
الجدول ٣: البيانات المصاحبة للمدخل المعجمي وفقًا لوثيقة دليل ترميز البيانات المعجمية والدلالية
(Lenci et al., 2000)

وتعد وثيقة منظمة الآيزو رقم ٢٤٦١٣ بعنوان: إطار نظري لتمثيل البيانات المعجمية (Lexical Markup Framework – LMF) - (٢٠٠٨-٢٠٢١) من أهم الوثائق الخاصة بمعايير تمثيل البيانات المعجمية العامة، وقد طورتها اللجنة الفنية لإدارة الموارد المعجمية واللغوية بالمنظمة (TC37). وصدرت نسخ متعددة من الوثيقة من (٢٠٠٣) وحتى آخر تحديث للوثيقة في (٢٠٢١)، ويتكون الإطار من أجزاء متعددة، ويستهدف

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً



النمـوذج المقـترح في الوثيقـة بالدرجـة الأولى المعاجـم الحاسـوبية المعـدة لغـرض الاسـتعمال في تطبيقـات المعالجـة الآليـة للغـة. ويعتمـد النمـوذج المعجـمي وفقًـا لهـذا الإطـار علـى لغـة النمذجـة الموحـدة (Unified Modeling Language – UML)، القابلـة للتحويـل إلى لغـة (XML) الـتي تسـتعمل في تمثيـل البيانـات اللغويـة، ويظهـر في (الشـكلين ٥ و٦) النمـوذج الأسـاسي لتمثيـل المعجـم الإلكـتروني. (Francopoulo, 2013).




الشكل ٥: النموذج الرئيس لنموذج (LMF) لتمثيل المعاجم المحوسبة.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

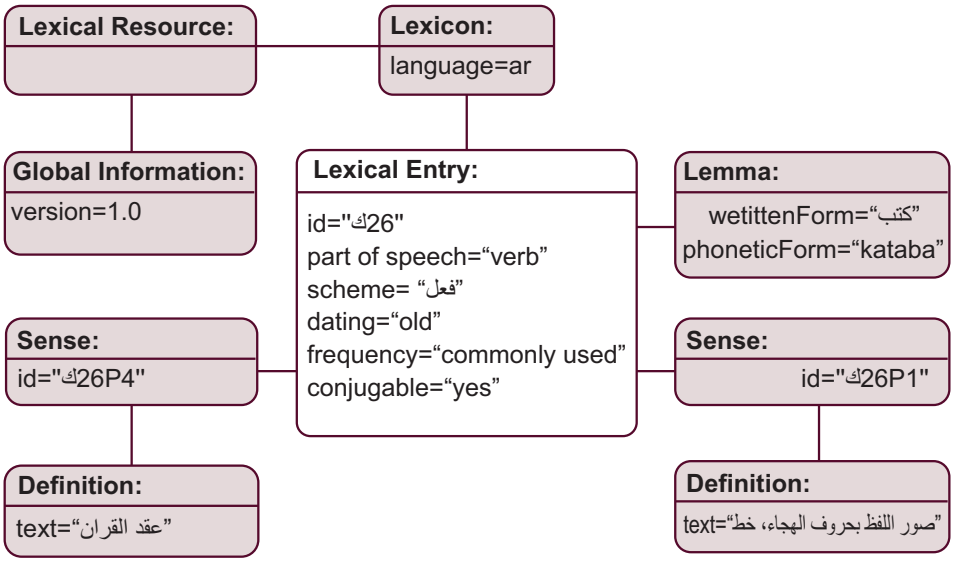
```
<LexicalResource dtdVersion="15">
  <GlobalInformation>
    <feat att="languageCoding" val="ISO 639-3"/>
  </GlobalInformation>
  <Lexicon>
    <feat att="language" val="eng"/>
    <LexicalEntry>
      <feat att="partOfSpeech" val="commonNoun"/>
      <Lemma>
        <feat att="writtenForm" val=" "/>
      </Lemma>
      <WordForm>
        <feat att="writtenForm" val=" "/>
        <feat att="grammaticalNumber" val="singular"/>
      </WordForm>
      <WordForm>
        <feat att="writtenForm" val=" "/>
        <feat att="grammaticalNumber" val="plural"/>
      </WordForm>
    </LexicalEntry>
  </Lexicon>
</LexicalResource>
```

الشكل ٦: نموذج (LMF) لتمثيل المداخل المعجمية في لغة (XML).

ويتكون النموذج من فرعين: الأول لتمثيل البيانات الأساسية للمداخل المعجمية، والآخر لتمثيل البيانات الإضافية؛ بما يتيح إدارة مشاركة المعجم وإعادة استعماله ودمجه مع المصادر اللغوية الأخرى. وقدم (Khemakhem et al., 2013) محاولة لنمذجة المعجم العربي المحوسب باستعمال (LMF). ويظهر (الشكل ٧) مثالا لتمثيل الفعل (كتب) وفقًا للنموذج.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً






الشكل ٧: تمثيل الفعل "كتب" وفقًا لنموذج (LMF) لتمثيل البيانات المعجمية.

ومن المعايير الأخرى المنشورة في تنظيم طرق عرض المداخل المعجمية وثيقة الآيزو ١٩٥١ لتمثيل وتقديم المداخل في المعاجم (٢٠٠٧-٢٠٢١). فقد نُشرت النسخة الأولى منها عام ١٩٧٣، وتضمنت صيغاً معيارية يمكن استعمالها في تمثيل بيانات المعاجم المطبوعة. وفي ١٩٩٧ صدرت النسخة الثانية التي اهتمت أيضاً بمعايير تنظيم عرض البيانات المعجمية المنشورة ورقيًا. ومع أن المعاجم الإلكترونية كانت قد بدأت حينذاك، فإن عددها كان محدودًا؛ لذلك لم تكن هناك حاجة ملحة لإعداد معايير لربط ومشاركة البيانات المعجمية الإلكترونية. وفي ٢٠٠٧ صدرت نسخة محدثة من وثيقة المعايير تضمنت معايير تفصيلية جديدة لتمثيل البيانات المعجمية العامة والخاصة، **وقد هدفت إلى:**

- دعم بناء وإدارة المعاجم الإلكترونية بأنواعها المختلفة.
- تسهيل إعادة استعمال ونشر المحتوى المعجمي في قوالب إلكترونية متنوعة.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

- تنظيم إجراءات إنتاج وتبادل وإدارة البيانات المعجمية.
- اقتراح نموذج لتصميم المعجم الإلكتروني يراعي أفضل الممارسات التطبيقية في صناعة المعاجم الإلكترونية (Derouin & Le Meur, 2008, p. 755).

وعلى الرغم من ريادة هذه الوثيقة المبكرة في معالجة المشكلة المعقدة الخاصة بتمثيل المعاجم الإلكترونية؛ فإن النموذج الأولي عانى من كثير من النواقص والسلبيات التي جعلت توظيفه غير ناجح في كثير من السياقات المعجمية (Lemnitzer, Romary & Witt, 2013). في حين راعت النسخة المحدثة من هذه الوثيقة معالجة أهم المشكلات في النموذج المعجمي السابق، واهتمت بتوظيف أحدث تقنيات لغة xml لتمثيل البيانات المعجمية، إضافة إلى الفصل بين التركيبي البنيوي للمعجم وواجهة المستخدم. وتعمل منظمة آيزو حالياً على إصدار نسخة محدثة جديدة لا تزال تحت الدراسة، ولم تُنشر حتى تاريخ إعداد هذه الوثيقة.

**ويمكن تقسيم مراحل بناء المعجم الإلكتروني العام إلى ما يلي:**

- المرحلة الأولى: مرحلة الإعداد والتحضير:
  - تحديد المستخدم والسياق المحتمل لاستعمال المعجم.
  - تحديد المعلومات المعجمية الموافقة لاحتياجات المستخدم.
  - اختيار الطرق المناسبة والمصادر الملائمة لجمع المادة المعجمية.
- المرحلة الثانية: مرحلة التصميم والتنفيذ والتطوير:
  - اختيار النموذج المعتمد للمعجم (الوظائف – العلاقات – أنواع البيانات – طرق الوصول).
  - تحديد فريق العمل (الفريق التنفيذي– المعجميون – اللغويون – التقنيون).
  - جمع البيانات المطلوبة للمعجم.
  - كتابة دليل التأليف ولوائح تنظيم العمل.

- كتابة الخطة التنفيذية وتحديد المهمات ومؤشرات الأداء ومواعيد التسليم.
- تنفيذ المشروع (الإدارة والمتابعة وضبط الجودة).
- التحضير لنشر المعجم في أحد أوعية النشر الإلكتروني.

— المرحلة الثالثة: ما بعد إعداد المعجم:

- التسويق وتقديم المعجم للفئات المستهدفة.
- متابعة استخدامات المعجم وتقويمها.
- متابعة مدى تلبية المعجم لاحتياجات المستخدمين وتحديثه ليكون أكثر موافقة لاحتياجات مستخدميه.

أما معايير تقييم المعاجم الإلكترونية فقد قسمها جاكسون (Jackson, 1996) إلى فئتين: معايير داخلية وخارجية. ويقصد بالمعايير الداخلية: التحقق من صدق ما يذكره مصممو المعجم ومحرروه من مزايا وخصائص، والتأكد من مدى توافرها في المعجم. وبعبارة أخرى يكون الهدف من هذه المعايير هو أن يحقق المعجم الأغراض التي أنشئ من أجلها، وأن يكون مطابقا للوصف والوظائف الخاصة به. أما بالنسبة إلى المعايير الخارجية فتشمل: معايير الوضوح اللغوي في شرح مكونات المعجم، والمعايير الخاصة بتصميم المعجم وأهم وظائفه، ومدى دقة وشمول مصادر بياناته ومعلوماته حول اللغة المستهدفة.

وبالرغم من الجهود المبذولة في العمل المعجمي العربي؛ فلا تزال الحاجة ماسة إلى وثائق ومعايير مرجعية تسهم في تحسين جودة المعاجم العربية ورفع كفاءتها؛ لتيسر على القارئ العربي استعمالها وتوظيفها في تطبيقات متنوعة. وقد ذكر الفهري (٢٠٢٢) عددًا من المآخذ على المعاجم العربية الحديثة المطبوعة والإلكترونية، **يمكن تلخيصها في النقاط التالية:**

— صعوبة الوصول إلى المعلومة المعجمية للمستخدمين من مختلف الفئات العمرية والثقافية.

- قصور المعاجم المعاصرة عن شمول المادة اللغوية المستعملة والممثلة للسان العربي.
- عدم مراعاة تمثيل التطور الدلالي لمعاني المداخل المعجمية، وتكرار المعاجم بعضها لبعض.
- ضعف كتابة التعاريف المعجمية، وعدم وضوحها ودقتها في الدلالة على المعاني اللغوية المستعملة للألفاظ العربية المعاصرة.
- المبالغة في استعمال المرادفات لشرح المعنى وعدم توظيف طرق شرح المعاني الأخرى.
- ضعف العناية بتأصيل الألفاظ والمعاني العربية، وربطها بالكلمات الأخرى المتصلة بها.
- ندرة الأمثلة الواقعية التي تبين المعنى المعاصر للألفاظ، وتكرار أمثلة تاريخية أو غير مناسبة.
- إهمال تمثيل عدد كبير من العلاقات الدلالية في المداخل المعجمية، مثل: الاشتراك اللفظي وتعدد المعاني وترابطها.

كل هذه التحديات في المعاجم العربية المعاصرة تُبرز أهمية إعداد الأطر المرجعية والمعايير لبناء وتقويم المعاجم الإلكترونية العربية العامة، وهو ما يسعى الجزء التالي إلى تقديمه.

## نطاق المعايير وحدودها

تقدم الوثيقة مجموعة من المعايير المهمة، التي تشمل جميع مراحل بناء المعاجم الإلكترونية لأغراض عامة، ولا يشمل نطاقُها المعاجمَ الاصطلاحية أو المعاجم المختصة. وتتضمن إطارًا عامًّا يمكن عن طريق تطبيقه بناء معجم إلكتروني يحقق مبادئ الجودة، ويكون ملائمًا للأهداف واحتياجات الفئات المستخدمة للمعجم.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

## قائمة المعايير[1]:

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| **أولا: معايير عامة لإعداد المعجم الي العام** | ١. تحديد المتطلبات الرئيسة لبناء المعجم الرقمي العام وتصميمه | أ. تحديد أهداف المعجم ونطاقه والفئات المستفيدة. | تحديد أهداف المعجم الرقمي، وصياغته صياغة واضحة، مثل: تسهيل تعلم اللغة أو تحسين ترجمة المصطلحات، مع تحديد نطاقه، وتحديد الفئات المستفيدة، مثل الطلاب، أو الباحثين، أو المتخصصين في مجال معين. | ظهور أهداف المعجم ونطاقه والفئات المستهدفة | مراجعة وثائق المعجم عن طريق قائمة تحقق. |
| | | ب. تحديد احتياجات وسمات الفئات المستفيدة. | تحليل احتياجات وخصائص الفئات المستهدفة من المستخدمين، مثل: مستوى اللغة، ونوعية المصطلحات المطلوبة، وأساليب التعلم المفضلة، لضمان توافق محتوى المعجم مع توقعاتهم، وخدمة أغراضهم التعليمية أو المهنية. | نسبة تغطية احتياجات وسمات الفئات المستهدفة في تصميم المعجم | تحليل احتياجات – مقابلات واستبانات مع الفئات المستفيدة |
| | | ج. تحديد القيمة المضافة للمعجم. | توضيح المشكلة أو التحدي الذي يهدف المعجم إلى حله، مثل: نقص الموارد اللغوية المتاحة بلغة معينة، أو صعوبة الوصول إلى معاجم متخصصة، أو الحاجة إلى تبسيط معاني المصطلحات لغير الناطقين باللغة. | نسبة وضوح المشكلة والفجوة اللغوية التي يعالجها المعجم | مراجعة الأهداف – تحليل فجوات لغوية – مقابلة خبراء |

(١) تم إدراج مؤشرات أداء متوسطة، لا تتطلب – غالبا – إجراءات مطولة، بحيث يمكن الإفادة منها بيسر وسهولة.

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | د. تصميم وظائف المعجم ومكوناته ومحتواه بناء على الأهداف واحتياجات فئات المستخدمين. | أن تُصمَّم مكونات المعجم –مثل: قائمة الكلمات، والتعريفات، والأمثلة، والنطق–، وتُطوَّر الوظائف –مثل: البحث المتقدم، أو التصنيفات، أو ميزات الترجمة، أو التوضيح التفاعلي؛ بناءً على الأهداف المحددة واحتياجات المستخدمين. | نسبة الوظائف والمكونات المطورة بما يلبي الأهداف والاحتياجات | اختبار قابلية الاستخدام – مراجعة تقنية – استبانة المستخدمين |
| | | هـ. إعداد دليل مرجعي لمستخدم المعجم العام. | إعداد دليل شامل يشرح كيفية استخدام المعجم، بما في ذلك أهداف المعجم، والفئات المستفيدة، وأساليب جمع البيانات، وكيفية استخدام الوظائف المختلفة، مثل: البحث، أو الحفظ، أو مشاركة النتائج؛ مما يسهل على المستخدمين الإفادة القصوى من إمكانات المعجم. | نسبة توفير الدليل للبيانات المطلوبة من المستخدمين | اختبار قابلية الاستخدام – مراجعة المعجم عن طريق قائمة التحقق. |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| **ثانيًا: معايير تجربة المستخدم** | ١. دعم المستخدم ومساعدته | أ. سهولة الوصول إلى بيانات الدعم والمساندة في المعجم. | توفير قنوات واضحة وسهلة للدعم والمساندة؛ ليتمكن المستخدمون من الوصول السريع إليها عند الحاجة للمساعدة في استخدام المعجم. | نسبة سهولة إمكانية الوصول إلى بيانات الدعم | اختبار فعلي لواجهة المعجم - استبانة للمستخدمين. |
| | | ب. وضوح بيانات المساعدة ومناسبتها لسياقات الدعم المختلفة. | أن تكون بيانات المساعدة واضحة ودقيقة ومناسبة للسياقات التي يحتاج فيها المستخدمون إلى دعم فني أو إرشادات حول استعمال المعجم. | نسبة وضوح وملاءمة بيانات المساعدة | تحليل محتوى - استبانة المستخدمين |
| | | ج. إضافة تدريب تفاعلي لطرق استعمال المعجم والبحث في بياناته. | أن تقدم منصة المعجم تدريبًا تفاعليًا يساعد المستخدمين على معرفة كيفية استخدام المعجم، بما في ذلك طرق البحث والاستعمال. | نسبة وجود وتفعيل وحدات التدريب التفاعلي | مراجعة برمجية - تتبع التفاعل داخل المنصة |
| | ٢. تخصيص تجربة المستخدم | أ. دعم الاحتياجات المخصصة للفئات المستفيدة من المعجم. | تكييف المعجم لتلبية احتياجات المستخدمين المختلفة، سواء طلابًا كانوا أم باحثين أم متخصصين، بتقديم محتوى وخدمات تتناسب مع احتياجاتهم التعليمية أو المهنية. | نسبة الخدمات المخصصة المقدمة للفئات المستهدفة | مراجعة التصميم - مقابلات مع الفئات المستفيدة |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ب. إمكانية إنشاء ملف تعريفي لمستخدم المعجم، تحفظ فيه أهم اختياراته وسجل استعماله للمعجم وتفضيلاته. | أن يُتاح للمستخدمين إنشاء ملف شخصي يحفظ تفضيلاتهم وسجل استخدامهم للمعجم؛ مما يسهل عليهم الوصول إلى المعلومات المفضلة بسرعة. | نسبة المستخدمين الذين تمكّنوا من إنشاء ملف تعريفي يحتفظ بتفضيلاتهم وسجل استخدامهم في المعجم. | تقارير النظام - تحليل بيانات الاستخدام |
| | | ج. إمكانية إعادة تصميم وتقديم محتوى المعجم وفقًا لتحليل الملف التعريفي للمستخدم. | تخصيص تجربة المستخدم بناءً على تحليل ملفه التعريفي، فيُعاد عرض المحتوى بطرق تناسب احتياجاته وتفضيلاته الشخصية. | نسبة نجاح إعادة تقديم المحتوى بناءً على الملف الشخصي | اختبار وظائف - استبيان رضا المستخدم |
| | ٣. تلبية تصميم واجهة المعجم احتياجات المستخدمين المتنوعة | أ. مرونة واجهة المستخدم ودعمها لبيئات وطرق عرض الأجهزة الإلكترونية المختلفة. | أن تكون واجهة المعجم مرنة وقابلة للتكيف مع مختلف الأجهزة، مثل: الهواتف الذكية والأجهزة اللوحية؛ لضمان تجربة سلسة عبر جميع المنصات. | نسبة توافق الواجهة مع مختلف الأجهزة | اختبار تقني متعدد الأجهزة |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ب. مناسبة التصميم لأهداف مستخدمي المعجم، ومراعاته لأغراض استعمالهم. | تصميم المعجم بما يتوافق مع أهداف المستخدمين، سواء أهداف تعلم اللغة أو البحث العلمي؛ لضمان تلبية احتياجاتهم بفعالية. | نسبة الرضا عن التصميم من قبل المستخدمين | استبانة خبراء - مقابلات المستخدمين |
| | | ج. الوضوح وسهولة الوصول إلى وظائف وبيانات المعجم. | أن تكون وظائف المعجم واضحة وسهلة الوصول؛ ليتمكن المستخدم من العثور السريع والسهل على المعلومات التي يحتاج إليها. | نسبة وضوح الوظائف وبيانات المعجم | استبانة المستخدمين. |
| | | د. تحكم المستخدم في وظائف المعجم وطريقة التنقل بين مكوناته المختلفة. | تطوير خيارات تحكم مرنة للمستخدم تتيح له التنقل بسهولة بين مكونات المعجم المختلفة، مثل: البحث عن الكلمات أو الوصول إلى التعريفات. | نسبة المرونة في التحكم والتنقل | اختبار عملي - مراجعة برمجية |
| | | هـ. تحقيق تطلعات المستخدم وتقديم المعلومات في القالب الإلكتروني المناسب لاحتياجاته. | أن يقدم المعجم المعلومات بما يلبي توقعات المستخدمين، مع واجهة إلكترونية تسهل الوصول إلى البيانات المطلوبة. | نسبة رضا المستخدم عن توافق المعلومات مع احتياجاته | استبانة مستخدمين - تحليل تعليقات |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | و. التعامل الأمثل مع أخطاء المستخدم المتوقعة أثناء تصفح المعجم المختص. | تصميم النظام ليكون قادرا على التعامل الفعال مع الأخطاء المتوقعة من المستخدم (مثل: البحث غير الصحيح)، ويتيح لهم التراجع عنها بسهولة وإعادة المحاولة. | نسبة معالجة الأخطاء الشائعة للمستخدمين | اختبار محاكاة أخطاء – مراجعة النظام |
| | | ز. تمكين المستخدم من تخصيص بعض الوظائف والبيانات وفقًا لاحتياجه وطبيعة استعماله للمعجم. | أن تقدم المنصة مرونة للمستخدمين لتخصيص تجربتهم، مثل: تعديل واجهة الاستخدام، أو ترتيب البيانات بطريقة تتناسب مع احتياجاتهم. | نسبة الوظائف القابلة للتخصيص | مراجعة برمجية – اختبار عملي |
| | | ح. البساطة في التصميم ووضوح التعليمات المقدمة للمستخدم بخصوص وظائف المعجم ومكوناته. | أن يكون تصميم المعجم بسيطًا وسهل الفهم، مع تعليمات واضحة تجعل من السهل على المستخدمين استخدام وظائف المعجم بدون تعقيد. | نسبة وضوح التصميم والتعليمات المقدمة للمستخدم. | استبيانات تقييم سهولة الاستخدام ووضوح التعليمات للمستخدمين |
| | | ط. التفاعل مع المستخدم وتقديم التغذية الراجعة الفورية والمناسبة. | أن يقدم المعجم تغذية راجعة فورية للمستخدمين أثناء تفاعلهم مع المحتوى؛ مما يساعدهم على التعلم من الأخطاء وتصحيحها بسرعة. | وجود خاصية توفير وتقديم التغذية الراجعة للمستخدمين. | اختبار عملي – تتبع النظام |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ي. مراعاة قدرات الذاكرة قصيرة المدى للمستخدم (عدم الجمع بين أكثر من خمسة عناصر تتطلب انتباه المستخدم في الوقت نفسه). | أن يأخذ المعجم في الاعتبار القدرات المحدودة للذاكرة قصيرة المدى للمستخدم، فلا تُعرض أكثر من خمسة عناصر تتطلب الانتباه في الوقت نفسه؛ لضمان سهولة الفهم. | نسبة تقليل الحمل المعرفي | مراجعة تصميم واجهة - تحليل تجارب المستخدم |
| | | ك. مراعاة التعددية الثقافية في تصميم واجهة المستخدم. | تصميم واجهة مستخدم تراعي التعددية الثقافية، لتكون مناسبة لمختلف الثقافات والخلفيات اللغوية التي قد تستخدم المعجم. | مستوى توافق التصميم مع معايير التعددية الثقافية | مراجعة خبراء - تحليل تصميم |
| | ٤. مراعاة سهولة الوصول والتنقل بين مكونات المعجم الإلكتروني العام | أ. سهولة التنقل بين مكونات المعجم المختلفة. | أن يكون التنقل بين مختلف أقسام المعجم (مثل: البحث، والفئات، والتعريفات) سلسًا وواضحًا؛ مما يتيح للمستخدم الوصول -بسهولة- إلى المعلومات المطلوبة. | نسبة نجاح المستخدمين في التنقل السلس | اختبار عملي - تتبع تجربة المستخدم |
| | | ب. كفاءة الوصول إلى البيانات المعجمية بأفضل الوسائل وأسرع الطرق. | أن يوظف المعجم طرقًا فعالة وسريعة للوصول إلى البيانات المطلوبة؛ مما يحسن تجربة المستخدم ويساعده في العثور على المعلومات بأقل جهد ممكن. | متوسط زمن الوصول إلى البيانات | اختبار أداء - تتبع الاستخدام |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ج. تقديم الدعم الملائم للمستخدم لاختيار الطريقة الأنسب في الوصول إلى البيانات. | تقديم إرشادات ودعم للمستخدمين لمساعدتهم على اختيار أنسب طريقة للوصول إلى المعلومات، سواء عبر البحث المباشر أو التصفح. | نسبة توفر طرق متعددة للوصول | مراجعة وظائف - استبانة المستخدم |
| | | د. التمييز بين روابط التنقل الداخلية والخارجية في المعجم. | تمييز الروابط الداخلية (التي توجه المستخدم داخل المعجم) عن الروابط الخارجية (التي توجه المستخدم إلى مصادر أخرى)؛ مما يساعد في تحسين تجربة التصفح. | نسبة وضوح الروابط الداخلية والخارجية | مراجعة تصميم - اختبار المستخدم |
| | | هـ. استخدام الروابط بطرق مناسبة والمساعدة على تنقل سلس ومنطقي بين مكونات المعجم. | استخدام الروابط استخداماً منطقيا ومترابطاً لتسهل التنقل بين مكونات المعجم المختلفة؛ مما يعزز تجربة المستخدم. | نسبة الروابط المستخدمة بشكل صحيح | مراجعة تقنية - تتبع تجربة المستخدم |
| | | و. استعمال الروابط التشعبية بكفاءة في داخل تعريف المدخل المعجمي وعلاقاته الدلالية. | أن تُستخدم الروابط التشعبية داخل التعريفات والمعاني بفعالية؛ مما يتيح للمستخدمين استكشاف العلاقات الدلالية بين الكلمات والمفاهيم بسهولة وتكامل. | نسبة التعريفات التي تحتوي روابط دلالية فعّالة | مراجعة محتوى - اختبار المستخدم |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| **ثالثًا: معايير المدخل المعجمي** | ١. تعريف المدخل المعجمي | أ. صياغة التعريف صياغة واضحة وسليمة لغويًّا، ومناسبة لفئات مستعملي المعجم. | كتابة التعريفات بأسلوب بسيط وسليم لغويًّا، بأن تكون مفهومة للفئات المستهدفة من مستخدمي المعجم. | نسبة التعريفات السليمة والواضحة | مراجعة لغوية - اختبار المستخدم |
| | | ب. شمول التعريف لجميع المعاني العامة للمدخل المعجمي. | يشمل التعريف كل المعاني العامة للكلمة أو المدخل في المعجم، ليحصل المستخدم على صورة شاملة عن استخدامات الكلمة في سياقات عامة مختلفة. | نسبة التعريفات الشاملة لجميع المعاني | مراجعة خبراء - مقارنة مع مصادر أخرى |
| | | ج. ارتباط المدخل المعجمي العام بمجموعة من المعاني المتقاربة المصنفة بحسب الدلالة أوالنوع الصرفي. | تصنيف المعاني المتقاربة للمدخل المعجمي وفقًا للدلالة أو النوع الصرفي؛ مما يسهل على المستخدم فهم العلاقات بين المعاني المختلفة للكلمة. | نسبة المداخل المصنفة دلاليًا أو صرفيًا | تحليل لغوي - مراجعة خبراء |
| | | د. التمييز بين المعاني المتقاربة والعلاقات الدلالية المختلفة، مع الحفاظ على اتساق في العرض والتصنيف لسهولة الفهم. | التفريق الواضح بين المعاني المتقاربة والعلاقات الدلالية المختلفة، مع الحفاظ على اتساق في العرض والتصنيف لسهولة الفهم. | نسبة التعريفات التي تميز بين المعاني المتقاربة | مراجعة محتوى - استبانة المستخدم |
| | | هـ. كتابة دليل منهجي لطريقة كتابة التعريفات المعجمية العامة. | إعداد دليل منهجي يوضح كيفية كتابة التعريفات المعجمية، بما في ذلك الإرشادات المتعلقة باللغة والأسلوب والتركيب، لضمان الاتساق والجودة في تحرير المعجم. | توفر دليل منهجي لصياغة التعريفات | مراجعة وثائق المنصة |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | و. مراعاة معيار الشيوع في ترتيب عرض الاستعمالات والمعاني المختلفة للمدخل المعجمي. | ترتيب المعاني المختلفة للكلمة وفقًا لشيوعها في الاستعمال، لتظهر الاستخدامات الأكثر شيوعًا أولاً؛ لتسهيل الوصول إليها. | نسبة المعاني المرتبة وفق معيار الشيوع | مراجعة المحتوى |
| | ٢. استيفاء المكونات الرئيسة في شرح المدخل المعجمي | أ. شمول المدخل المعجمي لفئات البيانات الرئيسة للمعجم العام (المدخل المعجمي — التعريف — التحليل الصرفي والنحوي — أمثلة الاستعمال التداولي). | أن يتضمن كل مدخل معجمي البيانات الأساسية، مثل: التعريف، والتحليل الصرفي والنحوي، وأمثلة الاستعمال؛ مما يُقدّم معلومات شاملة ومتكاملة للمستخدم. | نسبة المداخل المستوفية لجميع المكونات | مراجعة المحتوى – تحليل البيانات |
| | | ب. تضمّن الوحدة المعجمية العامة مجموعة من العلاقات الدلالية بالوحدات المعجمية الأخرى في المعجم (الترادف — التضاد — الاشتراك اللفظي). | ربط الوحدة المعجمية بعلاقات دلالية، مثل: الترادف، والتضاد، والاشتراك اللفظي؛ مما يساعد المستخدم في فهم العلاقات بين الكلمات. | نسبة المداخل التي تتضمن علاقات دلالية | مراجعة لغوية – مقارنة مع مصادر |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ج. تضمين فئات البيانات الإضافية المتصلة بأهداف المعجم واحتياجات مستعمليه، وفقًا للمعايير الدولية للمعاجم العامة في تحديدها. | تضمين بيانات إضافية تتوافق مع أهداف المعجم واحتياجات المستخدمين، مثل: التصنيفات اللغوية، أو الأمثلة السياقية، وفقًا للمعايير الدولية -مثل LMF- لتمثيل البيانات المعجمية. | نسبة المداخل التي تتضمن فئات | مراجعة لغوية - مقارنة مع مصادر |
| | | د. وضوح الأمثلة السياقية وأصالتها في الدلالة على المعنى (الاستعمال التداولي). | أن تكون الأمثلة السياقية واضحة وأصيلة، لتساعد في توضيح المعنى والاستخدام الفعلي للكلمة في سياقات مختلفة. | نسبة المداخل بأمثلة واضحة | مراجعة لغوية - مقارنة مع مصادر |
| | | هـ. استعمال تقنيات الوسائط المتعددة لإثراء المدخل المعجمي وتوضيح المعاني. | استخدام الوسائط المتعددة، مثل: الصور أو الفيديوهات؛ لتوضيح المعاني بصرياً، ولإثراء تجربة المستخدم. | نسبة المداخل التي تعرض وسائط | مراجعة محتوى |
| | | و. احتواء المدخل المعجمي على معلومات معرفية إثرائية عامة، مثل: الاقتباسات، والحكم والأمثال (الإثراء الثقافي). | أن يتضمن المعجم معلومات إضافية، مثل: الاقتباسات أو الأمثال التي ترتبط بالكلمة؛ مما يعزز الفهم الثقافي ويوسع المعرفة العامة للمستخدم. | نسبة المداخل التي تتضمن معلومات إثرائية | مراجعة المدخلات المعجمية |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ز. تضمّن المدخل المعجمي التراكيبَ والمتلازمات اللفظية ذات العلاقة. | تقديم المتلازمات اللفظية والتراكيب الشائعة مع الكلمة؛ مما يساعد المستخدمين على فهم كيفية استخدام الكلمة في سياقات مختلفة. | نسبة المداخل بالتراكيب ذات العلاقة | حصر التراكيب والمتلازمات |
| | | ح. توافر المقارنات المرجعية لمكونات المدخل المعجمي العام مع المصادر المناظرة محليًا وعالميًا. | إجراء مقارنات مرجعية بين مكونات المدخل المعجمي والمصادر المحلية أو العالمية الأخرى، لضمان دقة وشمولية البيانات مثل مقارنة المدخل المعجمي بنظيره في معاجم مشابهة. | نسبة المداخل ذات المقارنة المرجعية | مراجعة لغوية – مقارنة مع مصادر |
| | | ط. دعم التعدد اللغوي للمداخل المعجمية. | أن يدعم المعجم التعدد اللغوي بتقديم ترجمات وتعريفات لمداخل لغوية متعددة؛ مما يساعد المستخدمين ذوي الخلفيات اللغوية المختلفة. | نسبة اللغات المدعومة للمداخل | مراجعة المداخل المعجمية |
| رابعًا: المحتوى المعجمي | ١. جودة البيانات المعجمية | أ. مناسبة المحتوى لأهداف المعجم والفئات المستفيدة. | تصميم محتوى المعجم بما يتناسب مع أهدافه -سواء تعليمية كانت أم بحثية-، ومع احتياجات الفئات المستفيدة، مثل: الطلاب والباحثين. | نسبة مناسبة المحتوى للأهداف والفئات المستفيدة | تحليل المحتوى |
| | | ب. مناسبة مستوى التفاصيل وتعقيد المعلومات المعجمية المتصلة بالمدخل المعجمي. | أن تتناسب المعلومات المعجمية المقدمة مع مستوى الفئة المستهدفة، بأن تكون معقدة بما يكفي للمتخصصين، ومبسطة بما يناسب المتعلمين الجدد. | مستوى التفاصيل المتصلة بالمدخل المعجمي | تحليل المحتوى |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ج. اعتماد المعجم على مدونات لغوية تحتوي على بيانات ذات جودة عالية، مبنية وفق معايير بناء المدونات اللغوية؛ مما يضمن دقة المحتوى وتنوعه، وكذلك في الأمثلة السياقية المقدمة للمستخدمين. | أن يعتمد المعجم على مدونات لغوية تحتوي على بيانات ذات جودة عالية، مبنية وفق معايير بناء المدونات اللغوية؛ مما يضمن دقة المحتوى وتنوعه، وكذلك في الأمثلة السياقية المقدمة للمستخدمين. | نسبة المدونات عالية الجودة | مراجعة خبراء — تحليل المدونات |
| | | د. مناسبة لغة المدونة المستعملة في جمع بيانات المعجم، من حيث: النوع والمستوى اللغوي لأهداف المعجم، والفئات المستفيدة. | أن تتناسب لغة المدونة مع كلٍّ من:<br>- المستوى اللغوي المتوقع للفئة المستهدفة (رسمية- عامية- علمية) سواء مبتدئين كانوا أم متقدمين.<br>- الأهداف التعليمية أو البحثية للمعجم. | نسبة المدونات المتسقة | مراجعة خبراء — تحليل المحتوى |
| | | هـ. حداثة البيانات المعجمية، وتمثيلها للتطور الدلالي والاستعمال العام للغة. | أن تعكس البيانات المعجمية التطورات الحديثة في اللغة، وأن تكون محدثة بما يتلاءم مع الاستخدام الحالي والدلالي للكلمات. | نسبة المداخل الحديثة | مراجعة علمية - مقارنة بمصادر موثوقة |
| | | و. الربط بمصادر خارجية إثرائية متصلة بالمداخل المعجمية. | أن يحتوي المعجم على روابط لمصادر خارجية موثوقة تقدم معلومات إضافية متصلة بالمداخل المعجمية لتعزيز تجربة التعلم. | نسبة المداخل المربوطة | مراجعة لغوية - استبانة مستخدمين |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ز. التأكد من صدق مؤلفي البيانات المعجمية وموثوقيتهم. (المراجع) | أن يقدم المعجم معلومات حول مؤلفي البيانات المعجمية لضمان الموثوقية والمصداقية؛ مما يعزز ثقة المستخدمين في المحتوى. | نسبة المراجع الموثوقة | فحص المراجع |
| | | ح. تمكين المستخدم من تقييم المعلومات المعجمية والتعليق عليها. | يتمكن المستخدمون من تقديم تعليقاتهم على المعلومات المعجمية؛ مما يتيح التفاعل ويسهم في تحسين جودة المحتوى. | دعم التقييم والتعليق | مراجعة لغوية - تدقيق تقني |
| | | ط. وضوح اللغة وسلامتها من الأخطاء. | الحرص على أن تكون لغة المعجم واضحة وخالية من الأخطاء اللغوية والأسلوبية، أو أخطاء البيانات والمعلومات. | نسبة وضوح اللغة وسلامتها | مراجعة لغوية |
| | | ي. الابتعاد عن تضمين الاختصارات اللغوية غير الضرورية. | تجنب الاختصارات غير الضرورية التي قد تربك المستخدم، كالاختصارات التي تستعمل في المعاجم المطبوعة بسبب ضيق المساحة المتاحة. | نسبة الاختصارات في المداخل | مراجعة لغوية |
| | | ك. اتساق العلاقة ومناسبتها بين العناوين المستعملة والمحتوى الخاص بها. | تكون العناوين المستخدمة في المعجم متسقة مع المحتوى المرتبط بها؛ مما يضمن الوضوح وسهولة التنقل. | نسبة العناوين المتسقة مع المحتوى | تحليل المحتوى |
| | | ل. استعمال الوسائط المتعددة بفعالية لتوضيح المعاني والعلاقات المعجمية. | استخدام الوسائط المتعددة -مثل: الصور والفيديوهات- بفعالية؛ لتوضيح المعاني والعلاقات الدلالية بين الكلمات. | نسبة المداخل بوسائط | مراجعة نصوص - مراجعة الوسائط |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | م. توظيف المعلومات الإحصائية لتصنيف المداخل المعجمية بطرق مرئية (تصنيف دلالي- تصنيف بحسب الشيوع — تقنية سحابة الكلمات). | استخدام البيانات الإحصائية لتصنيف المداخل المعجمية بصور مرئية، مثل: سحابة الكلمات أو تصنيفات بحسب شيوع الاستخدام؛ مما يسهل على المستخدمين فهم الأنماط اللغوية. | نسبة البصريات الإحصائية إلى المداخل | مراجعة خبراء - تحليل محتوى |
| | ٢. مراعاة جودة تصميم وبناء نموذج البيانات المعجمية | أ. موافقة المعايير العالمية الملائمة في نمذجة البيانات المعجمية العامة. | أن يتبع المعجم المعايير العالمية في نمذجة البيانات -مثل LMF وغيرها- بحسب سياق المعجم؛ لضمان توافقها مع أنظمة وقواعد البيانات الأخرى. | نسبة البيانات الموافقة للمعايير | مراجعة خبراء |
| | | ب. توافق نموذج المعجم معأهداف مستخدميه واحتياجاتهم. | أن يتلاءم تصميم نموذج المعجم مع أهدافه واحتياجات المستخدمين، سواء أهداف التعلم أم البحث. | نسبة رضا المستخدمين | مراجعة محتوى - استبانة المستخدمين |
| | | ج. وضوح مكونات نموذج البيانات المعجمية والعلاقات والوظائف المرتبطة بها. | أن تكون مكونات نموذج البيانات المعجمية واضحة، مثل: المدخل المعجمي ومكوناته، مع تبيين العلاقات بين الكلمات والوظائف المختلفة التي يمكن للمستخدم الإفادة منها. | نسبة وضوح المكونات | مراجعة خبراء |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | د. شمول نموذج المعجم لكافة البيانات المعجمية التي يتضمنها المدخل المعجمي. | أن يتضمن نموذج المعجم جميع البيانات المعجمية ذات الصلة بالمدخل، بما في ذلك المعاني، والأمثلة، والتحليل الصرفي، والعلاقات الدلالية. | نسبة اكتمال النموذج | مراجعة محتوى |
| | | هـ. إمكانية مشاركة المعلومات المعجمية بطرق متعددة. | يقدم المعجم خيارات متعددة لمشاركة المعلومات، سواء عبر وسائل التواصل الاجتماعي أم عبر تصدير البيانات. | توفر طرق مشاركة | مراجعة تقنية – تحليل واجهة |
| | | و. سهولة إعادة استعمال بيانات المعجم في نماذج مشابهة. | أن تكون بيانات المعجم مرنة وسهلة لإعادة الاستخدام في مشاريع أو أنظمة معجمية أخرى؛ مما يزيد من فائدة المحتوى. | تمكين إعادة استخدام النموذج | مراجعة تقنية – تحليل واجهة |
| | | ز. إمكانية الإضافة والتعديل لمكونات نموذج تمثيل بيانات المعجم. | أن يتيح المعجم إمكانية إضافة البيانات المعجمية أو تعديلها، بما يساعد على تحديث المحتوى وتوسيعه بحسب الحاجة. | تمكين تعديل النموذج | مراجعة تقنية – تحليل واجهة |
| | | ح. تنظيم مكونات المدخل المعجمي العام تنظيماً منطقياً مناسباً لسياق استعمال المعجم. | أن تكون مكونات المدخل المعجمي منظمة تنظيماً منطقياً وسهل الفهم؛ مما يسهل على المستخدمين الوصول إلى المعلومات المطلوبة واستغلالها بفعالية. | تنظيم مكونات المدخل بطريقة مناسبة | مراجعة تقنية – تحليل واجهة |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| خامسـًا: معايير البحث في البيانات المعجمية وعرضها | ١. تصميم محرك البحث في المعجم الإلكتروني بطريقة تدعم تنوع خيارات المستخدم | أ. تصميم محرك البحث تصميماً ملائماً ومناسباً لسياق المعجم. | تصميم محرك البحث ليكون متكاملًا مع بنية المعجم. | متوسط زمن الاستجابة للبحث | اختبار أداء - تتبع النظام |
| | | ب. تصميم محرك البحث بطريقة تسهل التنقل بسلاسة بين صفحات المعجم المختلفة. | تصميم محرك البحث بطريقة تساعد المستخدم على الوصول إليه من أي جزء من المنصة، والتنقل بين صفحاته ومكوناته المختلفة بيسر وسهولة. | نسبة نجاح المستخدمين في التنقل السلس بين الصفحات | اختبار عملي - تتبع تجربة المستخدم - تحليل سجلات البحث |
| | | ج. توظيف سمات البحث المتقدم في المعجم عامةً، وفي محتوى المدخل المعجمي خاصةً. | يقدم محرك البحث خيارات متقدمة تتيح للمستخدمين توجيه البحث توجيهاً دقيقاً داخل محتوى المعجم، مثل البحث بالكلمات الرئيسية أو الفلاتر المتخصصة. | نسبة كفاءة استخدام خصائص البحث المتقدم | اختبار عملي - تتبع استخدام محرك البحث - استبانة رضا المستخدمين |
| | | د. توظيف تقنيات بحث متقدمة ويسيرة تحقق كفاءة وظائف البحث. | يتيح محرك البحث الوصول إلى المعلومات بسرعة وكفاءة، باستخدام تقنيات بحث متقدمة ويسيرة تضمن التحميل السريع للنتائج. | نسبة توفر تقنيات بحثية | تحليل الخدمات التقنية |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | هـ. تعدد الاختيارات والسمات التي يمكن استعمالها لتخصيص البحث، وتيسير الوصول إلى النتائج المستهدفة. | يمكن للمستخدمين تخصيص البحث باستخدام خيارات متعددة، مثل: الفلاتر، والمصطلحات اللغوية، والنوع الصرفي؛ مما يساعد على الوصول إلى النتائج الأكثر دقة. | نسبة تنوع وفعالية خيارات التخصيص في البحث | مراجعة وظائف محرك البحث - اختبار عملي مع المستخدمين - استبانة رضاهم |
| | | و. دعم طرق البحث عن الحروف، والرموز والكلمات المفردة والتراكيب. | يتيح محرك البحث إمكانية البحث باستخدام الحروف والرموز والكلمات المفردة أو التراكيب اللغوية؛ مما يتيح مرونة أكبر للمستخدمين في البحث. | توفر طرق البحث عن الحروف والرموز والكلمات المفردة والتراكيب. | اختبار عملي - مراجعة وظائف البحث - تحليل سجلات الاستخدام - استبانة المستخدمين |
| | | ز. إضافة ميزة الإكمال التلقائي للمدخلات المعجمية. | يتيح محرك البحث ميزة الإكمال التلقائي، بأن يقترح على المستخدم الكلمات الممكنة بناءً على الحروف التي يكتبها؛ مما يسهل الوصول إلى المدخلات. | وجود خاصية الإكمال التلقائي | مراجعة تقنية |
| | | ح. تقديم مقترحات مقاربة للمدخلات الخاطئة إملائيًا أو دلاليًا. | إذا كان المستخدم قد أخطأ في كتابة المدخل المعجمي، فيقدم محرك البحث مقترحات بديلة قريبة من الكلمة المكتوبة إملائيًا أو دلاليًا. | توفر التصحيح التلقائي | مراجعة تقنية — فحص أداء المنصة |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ط. تمكين التصحيح الإملائي الآلي للمداخل المعجمية، وإظهار البدائل المحتملة. | يقدم محرك البحث ميزة التصحيح الإملائي الآلي للكلمات المدخلة، مع عرض بدائل محتملة للمفردات التي قد يكون المستخدم أخطأ في كتابتها. | وجود التصحيح التلقائي للإملاء | مراجعة تقنية |
| | | ي. تمكين المستخدم من حفظ سجل البحث، وإعادة استعماله، وإضافته إلى المفضلة. | يمكن للمستخدمين حفظ سجل البحث الخاص بهم، وإعادة استخدامه لاحقًا، بالإضافة إلى إمكانية إضافة نتائج البحث إلى قائمة المفضلة. | وجود حفظ سجل البحث | مراجعة تقنية |
| | | ك. تمكين البحث الصوتي والكتابة الآلية للنصوص، مع تقديم التصحيح المناسب في حالة الخطأ في التعرف الآلي على الصوت. | يتيح المحرك خاصية البحث الصوتي، بأن ينطق المستخدمون كلمات البحث، مع تصحيح الأخطاء التي قد تحدث في التعرف الصوتي. | نسبة دقة البحث الصوتي مع التصحيح التلقائي للأخطاء | اختبار عملي للبحث الصوتي - تحليل سجلات النظام - استبانة رضا المستخدمين |
| | | ل. إظهار النوع الصرفي للمدخل المعجمي في قوائم الإكمال التلقائي لمحرك البحث. | عرض النوع الصرفي للكلمات المقترحة في ميزة الإكمال التلقائي؛ مما يساعد المستخدم على اختيار الكلمة المناسبة بناءً على التصنيف الصرفي. | نسبة المداخل التي يظهر معها النوع الصرفي في الإكمال التلقائي | مراجعة برمجية - اختبار عملي لمحرك البحث - استبانة رضا المستخدمين |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | م. تمكين تقنيات التعرف الآلي على الصور والنصوص المكتوبة. | يقدم محرك البحث إمكانية التعرف على النصوص المكتوبة في الصور؛ مما يتيح للمستخدمين البحث عن الكلمات أو العبارات عبر صور تحتوي على نصوص. | نسبة نجاح عمليات البحث باستخدام تقنيات التعرف على الصور والنصوص | اختبار عملي - تحليل دقة نظام التعرف - مراجعة برمجية |
| | | ن. تمكين البحث بالمعنى أو العلاقات الدلالية المختلفة للمدخل المعجمي. | يمكن للمستخدمين البحث عن كلمات أو عبارات بناءً على معانيها أو العلاقات الدلالية بينها وبين مفردات أخرى، مثل الترادف أو التضاد. | نسبة نجاح عمليات البحث الدلالي (ترادف، تضاد، اشتراك لفظي...) | اختبار عملي - مراجعة وظائف البحث - تحليل دقة العلاقات الدلالية |
| | | س. تمكين البحث بالقافية والسوابق واللواحق واللواصق الصرفية. | يدعم المحرك البحث بناءً على القوافي أو السوابق واللواحق الصرفية؛ مما يسهل إيجاد كلمات تتبع أنماطًا صرفية معينة. | نسبة نجاح عمليات البحث الصرفي (قافية / سوابق / لواحق / لواصق) | اختبار عملي - مراجعة وظائف البحث - تحليل مخرجات النظام - استبانة المستخدمين |
| | | ع. تمكين البحث بنمط الكلمة أو التركيب. | يمكن للمستخدمين البحث عن كلمات أو تراكيب وفقًا لأنماط معينة، مثل الجمل الاسمية أو الفعلية؛ مما يعزز من قدرة المستخدم على الوصول إلى نتائج دقيقة. | نسبة نجاح عمليات البحث بالنمط | اختبار عملي - مراجعة برمجية - تحليل مخرجات البحث |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ف. تمكين البحث بالأوزان الصرفية والنماذج الصوتية. | يتيح المحرك البحث باستخدام الأوزان الصرفية، مثل "فعَل" أو "مفعول"؛ مما يساعد في الوصول إلى كلمات تشترك في نفس النمط الصرفي. | نسبة نجاح عمليات البحث الصرفي/ الصوتي | اختبار عملي - مراجعة وظائف البحث - تحليل المخرجات |
| | | ص. تقديم مقترحات للمداخل المعجمية المشابهة أو المتصلة بعلاقات صرفية أو دلالية. | يعرض المحرك كلمات مشابهة أو متصلة بالمداخل المعجمية من حيث العلاقات الصرفية أو الدلالية؛ مما يسهل على المستخدم توسيع نطاق بحثه. | توفر خدمة تقديم المقترحات المشابهة | مراجعة وظائف النظام - اختبار عملي - استبانة رضا المستخدمين |
| | | ق. استخراج جميع المداخل المعجمية المتصلة، والتمييز بينها دون فتح الصفحات الخاصة بها. | يتمكن المستخدم من رؤية جميع المداخل المعجمية -ذات الصلة- مباشرة في نتائج البحث، دون الحاجة إلى فتح كل صفحة على حدة. | نسبة المداخل المتصلة المستخرجة بدقة | اختبار عملي - مراجعة وظائف البحث - تحليل استخدام فعلي |
| | | ر. تقديم عدد من الاختيارات الملائمة لاستعراض المعجم وفقًا لأنظمة ترتيب مخصصة . | يمكن للمستخدمين استعراض المعجم بطرق مختلفة، مثل: الترتيب الأبجدي أو حسب الشيوع؛ مما يجعل تصفح المحتوى متسما بمرونة عالية. | نسبة تفعيل أنظمة الترتيب المخصصة | مراجعة برمجية - اختبار عملي |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ش. توظيف المرشحات والفلاتر البحثية المتقدمة لتصفية نتائج البحث وإعادة تجميعها بطرق مختلفة. | يقدم المحرك فلاتر بحث متقدمة، مثل التصنيفات الصرفية أو الدلالية؛ مما يسمح للمستخدم بتصفية النتائج وإعادة تجميعها وفقًا لاحتياجاته. | نسبة كفاءة وفاعلية المرشحات في تحسين النتائج | اختبار عملي - مراجعة وظائف البحث |
| | | ت. عرض نتائج البحث بطرق مناسبة ومختصرة. | عرض نتائج البحث بطريقة مختصرة ومنظمة، فيحصل المستخدم على لمحة سريعة عن المعلومات دون الحاجة إلى التصفح المطول. | نسبة وضوح واختصار نتائج البحث | مراجعة محتوى - اختبار عملي |
| | | ث. توافر إمكانية التحكم والبحث في نتائج الاستعلام الخاصة بالمعجم. | يتمكن المستخدم من التحكم في نتائج البحث من خلال تعديل الاستعلامات أو تصفية النتائج للوصول السريع إلى المعلومات المطلوبة. | نسبة الوظائف المتاحة للتحكم في نتائج البحث | مراجعة برمجية - اختبار عملي |
| سادسًا: معايير الأدوات والتقنيات الإثرائية للمعجم | ١. توفير وظائف وأدوات توسع مجالات استعمال المعجم. | أ. إضافة مساعد ذكي لتعليم الكتابة. | يقدم المعجم أداة تعليمية ذكية تساعد المستخدمين على الكتابة الصحيحة للنصوص، مع تقديم ملاحظات حول الأخطاء الإملائية والنحوية. | توفر المساعد الذكي للكتابة | مراجعة برمجية - تتبع الاستخدام |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ب. إضافة أداة لتحليل النصوص المكتوبة وتصنيفها أو تمييزها آليًا وفق بيانات المعجم. | يتيح المعجم أداة لتحليل النصوص المكتوبة وتصنيفها تلقائيًا بناءً على البيانات المعجمية؛ مما يساعد المستخدمين في فهم النصوص وتفصيلها. | توفر أداة التحليل والتصنيف الآلي للنصوص | اختبار عملي - مراجعة نتائج التحليل - استبانة المستخدمين |
| | | جـ. إضافة أدوات مساعدة لاستعمال ذوي القدرات الخاصة للمعجم. | يدعم المعجم أدوات مساعدة لذوي القدرات الخاصة، مثل: قارئات الشاشة أو التكبير التلقائي؛ لضمان تجربة استخدام شاملة. | توفر وتفعيل أدوات المساعدة لذوي القدرات الخاصة | مراجعة برمجية - اختبار قابلية الوصول - استبانة ذوي القدرات الخاصة |
| | | د. توظيف تقنيات مبتكرة وحديثة لإدارة البيانات المعجمية وتصفحها والبحث فيها. | يعتمد المعجم على تقنيات مبتكرة لإدارة البيانات المعجمية؛ مما يضمن تجربة بحث سلسة وسريعة، مع إمكانية التنقل الفعّال بين المداخل. | نسبة تطبيق التقنيات الحديثة في إدارة البيانات والتصفح | مراجعة تقنية - اختبار عملي - تقارير النظام |
| | | هـ. استعمال أنظمة التوصية القائمة على الذكاء الاصطناعي لتقديم مقترحات المداخل المعجمية للمستخدم. | استخدام نماذج الذكاء الاصطناعي وأنظمة التوصية الذكية؛ لتقديم مقترحات للمستخدم بناءً على تاريخ بحثه أو استعلاماته السابقة؛ مما يعزز من تجربة الاستخدام. | نسبة دقة وملاءمة المقترحات المقدمة عبر أنظمة التوصية | مراجعة وظائف النظام - اختبار عملي - تحليل رضا المستخدمين |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | و. التوسيم اللغوي للنصوص والأمثلة المستعملة في المعجم (الصرفي - النحوي - الدلالي - التداولي). | تصنيف النصوص والأمثلة وفقًا لمعايير لغوية، مثل: التصنيف الصرفي والنحوي والدلالي؛ مما يسهل على المستخدمين فهم السياق اللغوي للمداخل. | نسبة النصوص الموسومة وفق المستويات اللغوية المحددة | مراجعة لغوية - تدقيق خبراء |
| | | ز. استعمال الرسوم البيانية والمشجرات لتنظيم بيانات المعجم وتيسير التنقل بين أجزائه. | تستخدم الرسوم البيانية والمشجرات لعرض العلاقات بين المداخل المعجمية بصورة مرئية؛ مما يسهل التنقل بين المفاهيم الدلالية والصرفية المختلفة. | نسبة المداخل التي تحتوي على رسوم بيانية أو مشجرات. | مراجعة تصميم |

## المراجع :


- الفهري، عبد القادر. (٢٠٢٢). المعجم العربي البنائي التنوعي؛ أسسه ونماذجه وقضاياه. عمان: دار كنوز المعرفة.

- Antoni-Lay, M. H., Francopoulo, G., & Zaysser, L. (1994). A generic model for reusable lexicons: The Genelex project. *Literary and Linguistic Computing, 9*(1), 47–54.

- Bergenholtz, H., & Bergenholtz, I. (2013). One database, four monofunctional dictionaries. *Hermes, Journal of Language and Communication in Business, 50,* 187–207.

- Francopoulo, G. (Ed.). (2013). *LMF lexical markup framework.* John Wiley & Sons.

- Granger, S., & Paquot, M. (Eds.). (2012). *Electronic lexicography*. Oxford University Press.

- International Organization for Standardization. (2007). ISO 1951:2007: *Presentation/ representation of entries in dictionaries—Requirements, recommendations and information*.

- Khatchadourian, H., Modiano, N., Heyer, G., & Waldhör, K. (1994). Use and importance of standard in electronic dictionaries: The compilation approach for lexical resources. *Literary and Linguistic Computing,* 9(1), 55–64. https://doi.org/10.1093/llc/9.1.55

- Khemakhem, A., Gargouri, B., Haddar, K., & Ben Hamadou, A. (2013). LMF for Arabic. In G. Francopoulo (Ed.), *LMF lexical markup framework*. John Wiley & Sons.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً


- Lenci, A., Bel, N., Busa, F., Calzolari, N., Gola, E., Monachini, M., Roventini, A., & Zampolli, A. (2000). SIMPLE: A general framework for the development of multilingual lexicons. *International Journal of Lexicography,* 13(4), 249–263.
- Salgado, A. (2022). *Terminological methods in lexicography: Conceptualising, organising, and encoding terms in general language dictionaries* [Doctoral dissertation, Universidade NOVA de Lisboa]. Repositório da Universidade NOVA de Lisboa. https://run.unl.pt/handle/10362137023/
- Sanfilippo, A., Calzolari, N., Ananiadou, S., Gaizauskas, R., Saint-Dizier, P., Vossen, P., & Zampolli, A. (1999). EAGLES LE34244-: *Preliminary recommendations on lexical semantic encoding* [Final report].
- Šimunić, D., & Pavić, I. (2020). *Standards and innovations in information technology and communications*. Springer. https://doi.org/10.10074-44417-030-3-978/
- Zajac, R. (1999). On some aspects of lexical standardization. *In Proceedings of the ACL SIGLEX Workshop on Standardizing Lexical Resources*.
- Zauner, A. (1902). Die romanischen Namen der Körperteile: Eine onomasiologische Studie. *Zeitschrift für romanische Philologie*, 27(s27), 40–182.

# الفصل الرابع

## معايير بناء المعاجم المختصة

- ملخص
- مقدمة
- مدخل تعريفي
- الدراسات السابقة
- نطاق المعايير وحدودها
- قائمة المعايير
- المراجع



هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

## ملخص

تقدم هذه الوثيقة المعايير الواجب مراعاتها في مختلف مراحل بناء المعجم الإلكتروني المختص وتطويره وتقويمه. وتبدأ الوثيقة بمقدمة وتمهيد ومدخل تعريفي لتوضيح مفهوم المعجم المختص وبيان أهميته وفوائد إعداد المعايير والوثائق المرجعية لبناء المعاجم الإلكترونية المختصة، ثم يستعرض الفصل مجموعة من الوثائق ذات العلاقة بمعايير المعاجم المختصة مع بيان علاقتها بالمعايير المقدمة في هذا الفصل. ثم يحدد الجزء اللاحق نطاق المعايير ومجالها، ويختتم الفصل باستعراض قائمة معايير المعجم المختص مصنفة في فئات متعددة.

## مقدمة

تُعد المعاجم المختصة من أهم أنواع المعاجم وأكثرها استعمالًا، ويمكن الإفادة منها بطرق مختلفة في تطبيقات متعددة. فإذا كان الهدف العام للمعاجم حفظ اللغة العامة وبيان معانيها وتوثيق استعمالاتها المختلفة، فإن هدف المعاجم المختصة توثيق اللغة المستعملة لأغراض خاصة، وجمع المفردات والمصطلحات الخاصة بمجال معرفي أو مهني خاص ومحدد. ويمكن تعريف الوحدة المعجمية المختصة أو الاصطلاحية أنها: المفردة أو التركيب اللغوي الذي يستعمل في الدلالة على مفهوم خاص بحقل معرفي أو مجال عملي محدد. وتشمل كلمات الاختصارات والألفاظ العلمية والمهنية والتسميات والألقاب الخاصة وغيرها.

وقد ازداد الاهتمام في السنوات الأخيرة بالمعاجم الإلكترونية المختصة أو معاجم المصطلحات، بتأثير مجموعة من العوامل، منها: الثورة المعرفية المصاحبة للاقتصاد المعرفي، وما أفرزته من تدفق ضخم للمصطلحات والمفردات ذات المفاهيم الخاصة في كل تفاصيل الحياة المعاصرة ومجالاتها المختلفة. إضافة إلى الحاجة الملحة لنقل المصطلحات الجديدة إلى اللغة العربية؛ لإثراء اللغة بألفاظ ومصطلحات واستعمالات لغوية جديدة، والمحافظة على استعمالها وفعاليتها في التواصل بين

المختصين في الحقول المعرفية والمهنية المستجدة، وتسهيل فهم اللغة الخاصة لغير المتخصصين. وتؤدي هذه المعاجم دورا مهما في تيسير الترجمة وكتابة البحوث والدراسات المتخصصة، ورفع مستوى فهم الألفاظ والمصطلحات المختصة في أوساط العامة والمتخصصين، وتسهيل طرق التواصل العالمي وتنمية الوعي والتعدد الثقافي (Trap-Jensen, 2018; L'Homme & Cormier, 2014).

وفي مجال المعالجة الآلية للغات البشرية، تكتسب المعاجم الاصطلاحية أهمية خاصة؛ فقد ثبتت فائدتها وفعاليتها في تحسين نتائج المعالجة الآلية للغة في تطبيقات متعددة، مثل دورها المحوري في تحسين نتائج محركات البحث، والترجمة الآلية، وفهرسة المستندات، والاستخراج والتلخيص الآلي للمصطلحات والكلمات المفتاحية، عن طريق التجارب التي وظفت المعاجم المختصة في مراحل مختلفة من العمل الحاسوبي.

وتواكب الوثيقة الحالية هذا الاهتمام؛ إذ تقدم معايير ينبغي مراعاتها في بناء المعاجم الإلكترونية المختصة وتطويرها وتقويمها؛ لرفع كفاءة إعدادها، وزيادة مستوى جودة بياناتها وطرق تنظيمها. وعلى الرغم من الجهود المبذولة؛ فإن المعاجم العامة والمختصة ما تزال بحاجة إلى مزيد من البحث والدراسة لاقتراح المناهج والأطر النظرية والمعايير الموضوعية التي تسهم في رفع مستوى جودة الصناعة المعجمية، وتحقيق أهدافها المختلفة (Romary & Wegstein, 2012). وأشار جاكسون (Jackson, 2002) إلى أن مجال نقد المعجم والوصول إلى معايير موضوعية لبنائه ما يزال بحاجة إلى مزيد من البحث والدراسة. وعلى الرغم من تعدد المعاجم الإلكترونية المتاحة فإن كثيرًا منها لا يعدو أن يكون نسخة إلكترونية من المعجم الورقي المطبوع، دون استثمار حقيقي ومناسب للميزات والإمكانات التقنية (Trap-Jensen, 2018; Bergenholtz, Nielsen & Tarp, 2009).

وتساعد مثل هذه المعايير على مواجهة التحديات المصطلحية التي تواجه اللغة العربية، ومن أهمها: التدفق الهائل للمصطلحات المختصة من اللغات العالمية ولا سيما الإنجليزية، وعدم انتشار البدائل العربية أو غياب الاتفاق على استعمالها بين المختصين، وكذلك تعدد الجهات الواضعة للمصطلحات وعدم اتفاقها على آلية لتوحيد المصطلح العربي؛ مما أدى إلى ازدواجية المصطلح العربي ورجوع كثير من المختصين إلى اللغات الأخرى لتفادي هذا الاختلاف والتشتت في المفاهيم

المختصة، وغيرذلك من التحديات التي تحتاج إلى تضافر جهود المتخصصين من مجالات متعددة، مثل: اللسانيات التطبيقية، والمعاجم، وعلم المصطلح، والتقنيات اللغوية، وفروع العلوم الحديثة المتنوعة. وتمثل هذه الوثيقة جزءًا من هذه الجهود التي تسهم في اقتراح مجموعة من المعايير المرجعية الواجب مراعاتها في إعداد المعاجم الإلكترونية المختصة وصناعتها.

## مدخل تعريفي

صنف القاسمي (٢٠١٩) الموضوعات المتصلة بالمصطلح إلى ثلاث فئات **كما يوضحها (الشكل١):**

- الفئة الأولى: ما يتعلق بعلم المصطلح ودراسة المفاهيم وكيفية الوصول إلى المصطلح المناسب، والقياس عليه وتعميم استعماله.
- الفئة الثانية: البحث في مجال المصطلح وتأصيله التاريخي والمدارس المختلفة لدراسة المصطلحات وتطورها.
- الفئة الثالثة: صناعة المصطلح، بداية من جمع المصطلحات ثم تحريرها ونشرها في المعاجم.

والفئة الثالثة هي الفئة الأقرب إلى مجال هذه الوثيقة التي تركز فيما يتعلق بصناعة المعجم الإلكتروني المختص الذي يهدف إلى جمع أنواع متعددة من المصطلحات والوحدات المعجمية المختصة لأغراض متعددة. فيشمل نطاق الوثيقة الحالية ذكر المعايير الواجب مراعاتها في بناء المعاجم الإلكترونية المختصة. ويقصد بالمعجم الإلكتروني المختص في هذا السياق: وعاء إلكتروني يهدف إلى جمع أي نوع من الوحدات المعجمية الخاصة، ويراد بها: الوحدة المعجمية المختصة التي تتضمن المفردة أو التركيب الاصطلاحي، وتدل على مفهوم تخصصي معين في حقل معرفي أو مجال مهني محدد.




الشكل ١: المجالات العامة لدراسة المصطلح، (القاسمي، ٢٠١٩).

**وتتميز الوحدة المعجمية المختصة بعدد من السمات، من أهمها:**

- **أحاديـة المفهـوم والدلالـة**، فعلـى النقيـض مـن المعاجـم العامـة الـتي يشـيع فيهـا تعـدد المعـاني للكلمـة أو التعبـير الواحـد، فـإن الأصـل في المعاجـم المختصـة أن تـدل الوحـدة المعجميـة علـى مفهـوم واحـد. وإن تعـددت المفاهيـم الـتي تـدل عليهـا، فيعامـل كل مفهـوم معاملـة الوحـدة المعجميـة المسـتقلة بذاتهـا.
- **معياريـة المفهـوم** الـتي تقتـضي تحقـق دلالـة مسـتقرة ومتفـق عليهـا في الاسـتعمال اللغـوي للمجتمـع المسـتهدف، وتخرجهـا هـذه الضوابـط مـن نطـاق الاسـتعمال اللغـوي العـام إلى الخـاص.
- وغالبًـا مـا تُسـتعمل الكلمـات والعبـارات العامـة نفسـها في الدلالـة علـى المعـاني الخاصـة أو الاصطلاحيـة، فتكـرر اللغـة المسـتعملة مـن حيـث الشـكل، ويختلـف مفهومهـا في المعجـم المختـص.

وتفـرق بعـض المراجـع (Ježek, 2016) بـين كلمـتي المعجـم (lexicon) والقامـوس (dictionary)، فيقصـد بـالأولى مفهومًـا عامًـا يتضمـن كل كلمـات اللغـة المسـتعملة، وبالثانيـة مفهومًـا خاصًـا يشـمل المعاجـم الـتي تجمـع أنواعًـا مـن مفـردات وتراكيـب اللغـة لأغـراض

متعـددة، وتسـتعمل كلمـة "قامـوس" أحيانًا في الدلالـة على هـذا المعنى. غيرأن هـذه الوثيقة لا تتبـنى هـذا التميـيز، بـل تُسـتعمل فيهـا كلمـة المعجـم للدلالـة علـى المفهومـين السـابقين.

ويتضمـن مفهـوم المعجـم المختـص أو الاصطلاحي أنواعًـا متعـددة مـن المصـادر اللغويـة المختصـة الـتي تحمـل أسـماء متعـددة، مثـل: القامـوس الخـاص، والموسـوعة المعرفيـة، وقواعـد أو بنـوك المصطلحـات، وقوائـم شـرح المصطلحـات، والمعاجـم الدلاليـة أو الأنطولوجيـا، وغيرهـا مـن الأوعيـة الإلكترونيـة الـتي تتضمـن أنواعًـا مـن الوحـدات المعجميـة المختصـة. ويوضـح الجـدول التـالي عـددًا مـن الفـروق بـين المعجـم العـام والمختـص (Albert, 2001).

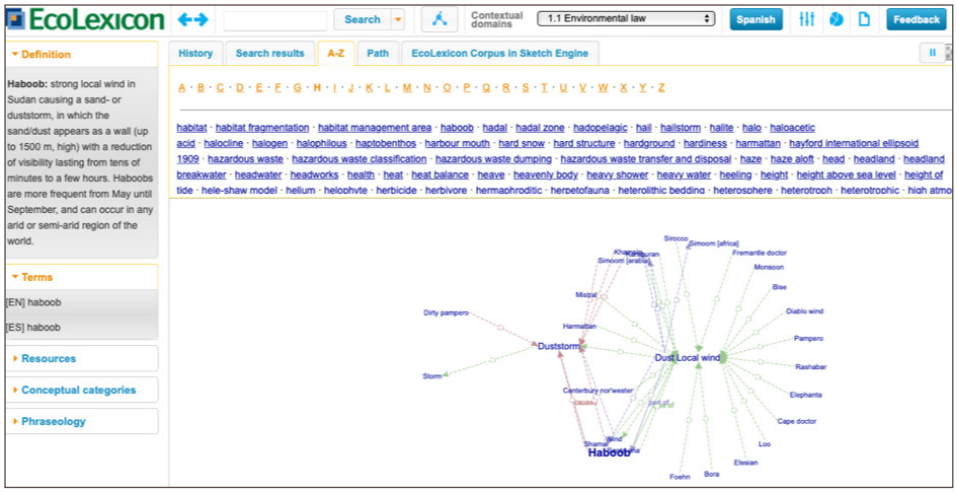
| م | وجه الفرق | المعجم العام | المعجم المختص أو الاصطلاحي |
|---|---|---|---|
| ١ | الهدف | جمع وتنظيم مفردات ومصطلحات اللغة العامة وبيان معانيها المتعددة. | جمـع وتنظيـم مفـردات ومصطلحـات مجـال معـرفي ومهـني محـدد، وبيـان معانيهـا الدقيقـة المختصـة. |
| ٢ | المحتوى والتنظيم | إعـداد معاجـم عامـة يكـون محتواهـا أوسـع، وأكـثر تنظيمـاً مـن المعاجـم المختصـة. | إعداد معاجم مختصة التي تركز في شرح المصطلحات المختصة بإيجاز، ووصف استعمالها في السياقات المختصة. |
| ٣ | رأس المدخل المعجمي | الوحـدة المعجميـة العامـة (كلمـة أو تركيـب)، وترتبـط غالبًـا بمعـان متعـددة وعلاقـات دلاليـة متنوعـة. | الوحـدة المعجميـة المختصـة (كلمـة أو تركيـب)، وترتبـط غالبًـا بمفهـوم واحـد في مجـال معـين. |
| ٤ | منهج التعريف | تعتمـد منهـج الوصـف والتوضيـح في تعريـف واسـتعمال المدخـل المعجـمي العـام لأغـراض تواصليـة في سـياق عـام.<br>(تصف الاستعمال الطبيعي العام). | تعتمـد منهـج الإرشـاد والتوجيـه في تعريـف واسـتعمال المدخـل المعجـمي الخـاص لأغـراض تواصليـة في سـياق تخصـصي محـدد.<br>(تصف الاستعمال الاصطناعي الخاص). |
| ٥ | طريقة الترتيب | المعجـم التقليـدي العـام يعتمـد غالبًـا علـى الترتيـب الألفبـائي. | المعجـم التقليـدي المختـص يعتمـد غالبًـا علـى المعـاني والحقـول الدلاليـة. |

| م | وجه الفرق | المعجم العام | المعجم المختص أو الاصطلاحي |
|---|---|---|---|
| ٦ | هيكلة المواد | تبدأ من تحديد الكلمات ثم البحث عن معانيها في الخطاب اللغوي العام. | تبدأ من تحديد المفاهيم ثم البحث عن الكلمات التي تستعمل في الدلالة عليها في الخطاب اللغوي الخاص. |
| ٧ | نوع اللغة التي تهتم بها | تهتم بدراسة اللغة الشفوية والمكتوبة. | تهتم غالبًا بدراسة اللغة المكتوبة لقلة استعمالها في اللغة الشفوية. |
| ٨ | المفردات المتضمَّنة | يتضمن المفردات والعبارات الوظيفية والفئات الصرفية المتعددة (اسم - فعل -حرف -صفة). | يتضمن المفردات والعبارات التي تشير إلى المحتوى المختص، ويركز غالبًا في الأسماء المفردة، وتشيع فيه التراكيب الاسمية الوصفية والاصطلاحية. |
| ٩ | جهة الإعداد | يُعده اللغويون، ويستفيد منه معظم فئات الناطقين باللغة. | يُعدّ محتواه المختصون في مجال محدد، بالتعاون مع المعجميين واللغويين. |
| ١٠ | الفئة المستهدفة | تُعد لاستعمال العامة لأغراض تعليم اللغة وتعلمها وتوثيقها. | تُعد لاستعمال المختصين في مجال معرفي أو مهني محدد، أو للمهتمين بصناعة المعاجم المختصة. |

الجدول ١: أهم الفروق بين المعاجم العامة والمختصة.

ومن أبرز الأمثلة على المعاجم الإلكترونية المختصة: المعجم البيئي الذي يتضمن ٤٦٧٧ مدخلًا معجميًا متصلة بموضوع البيئة، ويدعم هذا المعجم المختص -دعماً كاملاً أو جزئياً- ست لغات من بينها اللغة العربية. ويعتمد المعجم نظرية التمثيل الارتباطي للمداخل المعجمية من خلال ربطها بشبكة من العلاقات المنطقية تتجاوز أكثر من ٨٧٩٧ علاقة، ويبلغ معدل العلاقات لكل مدخل معجمي ٤ علاقات تقريبًا. ويوضح الجدول ٢ بعض أمثلتها.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً




| | | | | | |
|---|---|---|---|---|---|
| ١ | نوع من | ٤ | نتيجة ل | ٧ | يسبب |
| ٢ | يتأثر ب | ٥ | يقع في | ٨ | يضاد |
| ٣ | يؤثر في | ٦ | مصنوع من | ٩ | جزء من |

الجدول ٢: أمثلة للعلاقات في المعجم البيئي المختص (https://ecolexicon.ugr.es).

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً



## الدراسات السابقة

تشـترك المعاجـم المختصـة في كثيـر مـن مراحـل بنائهـا مـع المعاجـم الإلكترونيـة العامـة، لذلـك يسـتعرض هـذا القسـم عـددًا مـن الوثائـق والأطـر الـتي اهتمـت بوضـع معايـير عامـة لصناعـة المعاجـم، ولا سـيما الدراسـات الـتي ركـزت في معايـير المعاجـم الإلكترونيـة المختصـة، وتشـمل مراجعـة الوثائـق ذات العلاقـة تقديـم نبـذة عنهـا، وذكـر لأهـم المعايـير الـتي اشـتملت عليهـا، مـع الإشـارة إلى مـا يناسـب الوثيقـة الحاليـة مـن معايـير المعاجـم المختصـة.

وعنـد النظـر في معايـير المعاجـم المختصـة أو المصطلحـات الـتي أصدرتهـا منظمـة الآيـزو، نجـد مجموعـة مـن الوثائـق الـتي عالجـت جوانـب مختلفـة مـن صناعـة معاجـم المصطلحـات وبنائهـا وتخزينهـا إلكترونيـا وفـق معايـير مختلفـة[1]، **ركـزت -إجمـالاً- في:**

١- معايير تعريف المصطلحات وربطها بطرق متعددة.

٢- تنظيم وتحليل الحقول الدلالية للمعاجم المختصة.

٣- معايير تصميم وبناء الأنظمة الحاسوبية الخاصة بمعالجة المداخل المعجمية المختصة.

٤- معايير الترميز اللغوي وغير اللغوي للمصطلحات.

٥- معايير تنظيم وإنتاج وصيانة قواعد بيانات المعاجم المختصة.

وتُعـد معايـير آيـزو لإدارة قواعـد بيانـات المصطلحـات والمعاجـم المختصـة (ISO 30042) مـن أهـم الوثائـق المتعلقـة بمعايـير تمثيـل بيانـات المعاجـم الاصطلاحيـة الإلكترونيـة، فقـد بُنيـت علـى مجموعـة كبـيرة مـن الدراسـات والبحـوث المسـحية الخاصـة بمعرفـة طـرق إدارة ومعالجـة بيانـات المعاجـم الاصطلاحيـة، ونشـرت أول نسـخة منهـا في (٢٠٠٩) وتوالـت

(١) أصدرت منظمة الآيزو ممثلة في اللجنة الفنية المختصة بمعايير المعاجم والموارد الاصطلاحية منذ عام ١٩٨٥ أكثر من ٩ وثائق معايير عالمية لإدارة وجمع ومعالجة المصطلحات والمعاجم المختصة، انظر لمزيد من التفاصيل: https://www.iso.org/committee/48136/x/catalogue

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

التحديثـات المسـتمرة عليهـا حـتى آخـر إصـدار لهـا عـام (٢٠١٩). وعرفـت الوثيقـة المعجـم المختـص بأنـه: قاعـدة بيانـات تجمـع فيهـا مجموعـة مـن الكلمـات والمصطلحـات والمفاهيـم الـتي تسـتعمل لأغـراض خاصـة، وتتضمـن مداخلهـا المعجميـة عـدداً مـن التعريفـات والمرادفـات والعلاقـات الدلاليـة المتنوعـة. واعتمـدت وثيقـة المعايـير على منهجيـة العلاقـة الدلاليـة المنفـردة لـكل وحـدة معجميـة خاصـة، وذلـك عـن طريـق ربـط كل مدخـل معجـمي بمعـنى ومفهـوم اصطـلاحي خـاص، وفي حالـة تعـدد معـاني المدخـل فإنهـا تعتـبر مداخـل معجميـة مسـتقلة، وهـذا مخالـف لظاهـرة تعـدد المعـاني المعتـادة في المداخـل المعجميـة العامـة. ومـن المعايـير الـتي اعتمدتهـا الوثيقـة ويمكـن تبنيهـا في الوثيقـة الحاليـة، مجموعـة مـن المعايـير لضبـط كتابـة تعريـف المصطلـح الخـاص.

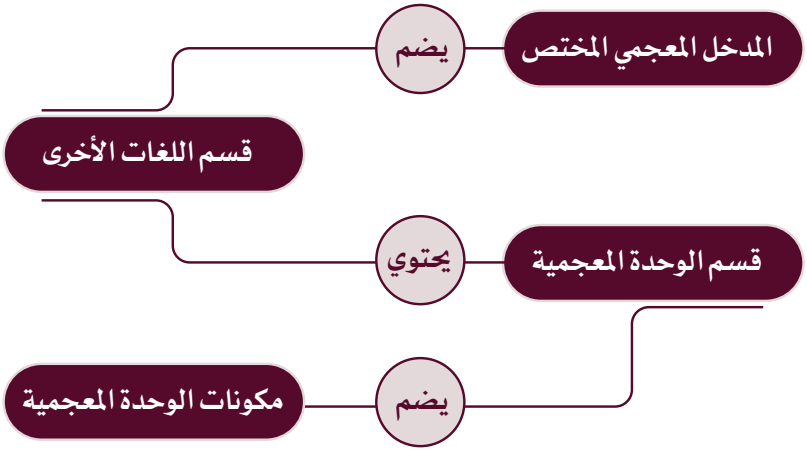
وتمثّـل الدافـع الرئيـس لهـذه الوثيقـة في تيسـير عمليـة تبـادل البيانـات المعجميـة الإلكترونيـة الخاصـة بـين اللغـات والتخصصـات المختلفـة، عـبر اعتمـاد معايـير عالميـة موحـدة لتنظيمهـا وتخزينهـا ومعالجتهـا في قواعـد البيانـات المحوسـبة. ووفقًـا لمعايـير الوثيقـة؛ فـإن معجـم المصطلحـات يتكـون مـن نوعـين مـن فئـات البيانـات: الفئـات الإلزاميـة (مثـل: المصطلـح أو رأس المدخـل المعجـمي، واللغـة، والتعريـف)، والفئـات الاختياريـة (مثـل: التحليـل الصـرفي، والحقـل الـدلالي، والتعليقـات). وأوردت الوثيقـة التمثيـل المناسـب لـكل هـذه البيانـات في لغـة xml المسـتعملة لتمثيـل بيانـات المـوارد اللغويـة، ويوضـح (الشـكل ٢) مثـالًا لفئـة الحقـل الـدلالي.

| 6.17 Subject field | |
|---|---|
| **Identifier** | www.isocat.org/datcat/DC-489 |
| **XML representation** | `<descrip type="subjectField">` |
| **Content type** | plainText |
| **Level** | Concept |
| **Comment** | It is best practice to implement this data category as a picklist. You can specify the picklist values in a multi-level hierarchy. The way you define the hierarchy depends on the terminology application that you use. |

الشكل ٢: مثال لتمثيل بيانات المعجم الاصطلاحي في فئة الحقل الدلالي للمدخل المعجمي.

وقدمـت وثيقـة أخـرى لمنظمـة الآيـزو (ISO 16642) إطـارًا مرجعيًـا لنمذجـة البيانـات المعجميـة المختصـة، تتضمـن أدوات ونمـاذج حاسـوبية لتمثيـل البيانـات المعجميـة متعـددة اللغـات، وتضمنـت الوثيقـة نموذجًـا رئيسًـا لتمثيـل المدخـل المعجـمي المختـص مـع مجموعـة مـن المعلومـات الإضافيـة الملحقـة بـه، مـع عـدد مـن فئـات البيانـات الـتي يمكـن أن يتضمنهـا المعجـم الإلكـتروني المختـص. ويقـوم النمـوذج علـى البنـاء الهـرمي الـذي يبـدأ بالإطـار العـام، ويتضمـن جميـع المداخـل المعجميـة المختصـة، ويشـمل كل مدخـل معجـمي مجموعـة مـن البيانـات ومـن ضمنهـا فئـة خاصـة بترجمـة المصطلـح في اللغـات الأخـرى. ويوضـح (الشـكل ٣) البنيـة الرئيسـة لنمـوذج تمثيـل البيانـات المعجميـة وفقًـا للمعيـار، ويحتـوي كل قسـم علـى عـدد مـن السـمات، مثـل: الرمـز التعريفـي للغـة، والوحـدة المعجميـة، والتعريـف، وغيرهـا.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً




الشكل ٣: النموذج الرئيس للمعجم الاصطلاحي المختص وفقا لمعيار آيزو ١٦٦٤٢.

ويحـدد معيـار (ISO 704:2022) المبـادئ الرئيسـة والمنهجيـات لتحليـل المفاهيـم والمصطلحـات وتوحيدهـا. ويضـع معايـير تسـاعد علـى صياغـة تعريفـات دقيقـة وواضحـة للمداخـل المعجميـة المختصـة، وربطهـا بالمفاهيـم بنـاءً علـى فهـم واضـح لسـياقها واسـتخدامها. إضافـة إلى منهجيـات لتحديـد المصطلحـات وحـل التناقضـات أو التداخـلات فيهـا؛ لضمـان احتـواء المعجـم المختـص علـى مداخـل موحـدة ومتسـقة. كمـا قدمـت الوثيقـة إرشـادات لتحليـل المفاهيـم، وطـرق تجميـع المصطلحـات المتشـابهة منطقيًـا بنـاءً علـى علاقاتهـا المفاهيميـة.

وتقـدم الوثيقـة إطـارًا مرجعيًـا ومجموعـة مـن المعايـير لتمثيـل البيانـات المتصلـة بالمعاجـم المختصـة وتبادلهـا، واسـتعملت لغـة XML الحاسـوبية في ترميـز البيانـات المعجميـة، وتتضمـن البيانـات الخاصـة بالمعجـم ٢١ فئـة وفقًـا للنمـوذج، يوضحهـا الجـدول ٣.

| التاريخ | ملحوظات | المدخـــل المعجـمي | حالـــة التحديـث | النـوع الصـرفي للمصطلـح | الفئة الفرعية | الحقـــل الـدلالي |
|---|---|---|---|---|---|---|
| رابط خارجي | أمثلـــة سـياقية | التعريف | الجنـــس النحـوي | استعـمـال مناطقـي | مشـــروع فرعـي | مصطلحـا ت ذات علاقـة |
| مفاهيـم ذات علاقـة | المسؤولية | نوع المصطلح | نـوع التمثيـل الكتـابي | شبكـــة المفاهيـم | موقـع المدخـل المعجـمي | المصدر |

الجدول ٣: فئات البيانات وفقًا لمعاييرآيزو ٣٠٠٤٢ لتمثيل البيانات المعجمية المختصة.

وتصنـف هـذه الفئـات مـن حيـث أهميـة تضمنهـا في المعجـم إلى فئـات رئيسـة ومتوسـطة وشـاملة، وتشـمل الفئـات الرئيسـة الـتي ينبغـي أن يشـتمل عليهـا أي معجـم مختـص: (المدخـل المعجـمي، والتاريـخ، والملحوظـات)، وتتضمـن الفئـة الـتي تليهـا إضافـة إلى مـا سـبق: (حالـة التحديـث، والفئـة الفرعيـة، والنـوع الصـرفي للمدخـل المعجـمي، والحقـل الـدلالي، والرابـط المرجعـي الخـارجي). وروعـي في تصميـم النمـوذج أن يكـون مرنًـا ليتضمـن كل المعلومـات المتعلقـة بالمدخـل المعجـمي المختـص، ويمثـل التنـوع الواسـع في بيانـات المعجـم. ومـن أهـم المهمـات الـتي يسـعى النمـوذج لتحقيقهـا:

- المسـاعدة علـى حفـظ وتوثيـق بيانـات الوحـدات المعجميـة المختصـة، واسـتعمال طريقـة الحفـظ المسـتقل لمحتـوى المعجـم المنفصـل عـن أي أداة أو برنامـج حاسـوبي معـين، ليمكـن الإفـادة منهـا في منصـات وتطبيقـات حاسـوبية متعـددة.
- تيسـيرتبـادل البيانـات المعجميـة المختصـة وإعادة اسـتعمالها لأغـراض متنوعـة، ولا سـيما بـين التطبيقـات المتصلـة بتأليـف المعاجـم المختصـة وترجمتهـا واسـتخراجها آليًـا مـن المدونـات اللغويـة.
- تقديـم نمـوذج جاهـز لتطويـر المعاجـم المختصـة يسـهل مـن إجـراءات جمـع المداخـل المعجميـة المختصـة ونشـرها، بأفضـل الأنظمـة والتصاميـم والأدوات الحاسـوبية.

وتعد وثيقة آيزو (ISO 12620) إحدى وثائق المعايير المعتمدة لنمذجة وإدارة البيانات المعجمية المختصة، وهي تقدم وصفًا تفصيليًّا لفئات البيانات التي يمكن أن يحتوي عليها المعجم المختص، ويمكن تصنيف تلك البيانات إلى بيانات خاصة بمفاهيم الوحدات المعجمية المختصة -وتكون في الغالب مشتركة بين اللغات-، وبيانات أخرى خاصة بلغة معينة، وبيانات خاصة بوحدة معجمية معينة. ويوضح (الجدول ٤) بعض الأمثلة على فئات البيانات التي تضمنها هذا المعيار.

| فئة البيانات | أمثلة البيانات المعجمية |
|---|---|
| التحليل الصرفي النحوي | النوع الصرفي، والجنس، والعدد، والمشتقات والجذور. |
| الصوتيات | النطق الصوتي، والكتابة الصوتية. |
| التأصيل التاريخي | التطور الدلالي، والأصل اللغوي. |
| الوحدات ذات العلاقة | طرق الكتابة البديلة، والاختصارات، والبدائل الأخرى. |
| التحليل الدلالي | التعريف، والتحليل الدلالي، والعلاقات الدلالية. |
| التراكيب والمتلازمات اللفظية | الألفاظ اللغوية المصاحبة للمدخل المعجمي. |
| الاستعمال التداولي | أمثلة سياقية لاستعمال المدخل المعجمي. |
| الرموز التعريفية | الرموز التي تشير إلى المصطلح في قواعد البيانات ذات الصلة. |
| الحقل الدلالي والمعرفي | الحقول الدلالية العامة والفرعية. |
| المصدر والروابط التشعبية | المراجع الإضافية، والمصادر ذات العلاقة، الروابط، ...إلخ. |

الجدول ٤: أمثلة لمجموعات البيانات التي يتضمنها معيار آيزو ١٢٦٢٠ لإدارة البيانات المعجمية المختصة.

وتشمل صناعة المعجم الإلكتروني المختص خمس مراحل، **وهي -باختصار- كما يلي:**

**١- التمهيد والتحضير (تحديد الفئة المستهدفة وأهداف المعجم ونطاقه)، وتشمل:**

- تحديد الموضوع العام للمعجم المختص.
- تحديد الفئة المستهدفة ودراسة أهم احتياجاتها.
- كتابة الأهداف العامة للمعجم.
- تحديد نطاق المعجم وسياقات استعماله المحتملة.

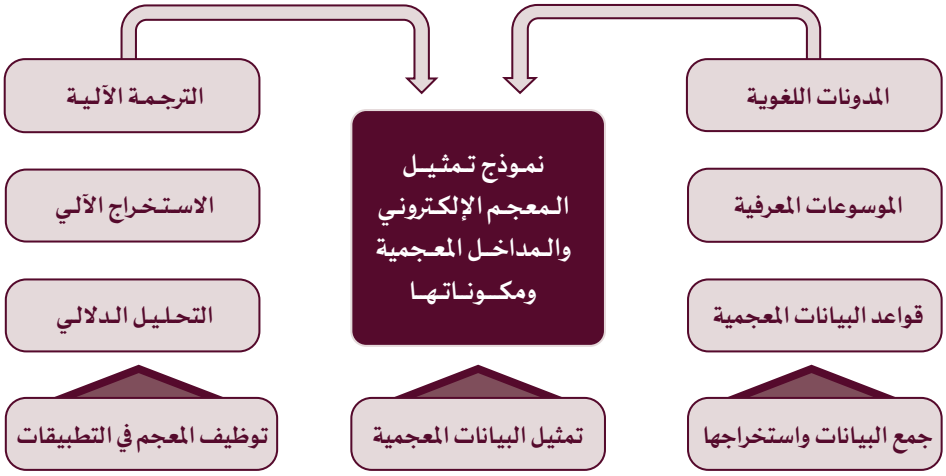
**٢- جمع المصادر اللغوية (بناء المدونة اللغوية المختصة)، وتشمل:**

- تحضير البيانات وتنسيقها وفقًا لأهداف المعجم.
- تصنيف النصوص وفقًا للموضوعات الرئيسة والفرعية.
- ترميز أو توسيم النصوص لغويًا ودلاليًا (من الوسوم العامة المقترحة: المنطقة - معلومات حول سياق النص - طبيعة المحتوى - المستوى اللغوي - نوع اللغة مسموعة أو مكتوبة - توقيت جمع النص أو مصدر كتابته).
- ضمان معايير تمثيل المدونة للنوع اللغوي المستهدف.

**٣- تصميم واختيار نموذج تمثيل البيانات المعجمية، ويراعى فيها:**

- أن يكون نموذج البيانات موافقًا للمعايير العالمية في تمثيل البيانات المعجمية لتسهيل نقل البيانات المعجمية، وإعادة استعمالها وربطها بالمصادر المعجمية الأخرى.
- اتساق نموذج البيانات مع أهداف المعجم والفئات المستهدفة. ويوضح (الشكل ٤) بعض مكونات نماذج البيانات المعجمية في جمع البيانات، وتمثيلها، والتطبيقات العملية للمعجم.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً




الشكل ٤: أمثلة لمكونات نموذج تمثيل البيانات المعجمية.

٤- استخراج وجمع المصطلحات والمداخل المعجمية، وفي هذه المرحلة تستعمل المناهج الحاسوبية الملائمة لاستخراج المداخل المعجمية المختصة (نموذج الاستخراج المعرفي أو المبني على البيانات والتحليل الإحصائي للمدونة أو النماذج المختلطة).

٥- إضافة التعاريف وإكمال المعلومات الملحقة بالمداخل المعجمية، وتشمل هذه المرحلة:

- إكمال بيانات المداخل المعجمية وفقًا للنموذج الحاسوبي المعتمد.
- مراعاة الدقة والمرونة في تمثيل البيانات، مع تيسير ربطها ونقلها بين الأنظمة الحاسوبية المختلفة.
- مراعاة سهولة الوصول إلى البيانات وعلاقتها بأهداف المعجم واحتياجات الفئات المستهدفة.

ويوضح (الجدول ٥) أمثلة لمكونات المدخل المعجمي التي اشتملت عليها معظم الوثائق السابقة الخاصة بمعايير بناء بيانات المعاجم المختصة وتمثيلها.

| مكونات المدخل المعجمي المختص | |
|---|---|
| عنوان الوحدة المعجمية المختصة | المصدر |
| نوع الوحدة المعجمية المختصة | الأمثلة السياقية |
| الرمز التعريفي | وحدات ذات علاقة |
| الرمز اللغوي | مجال التطبيق |
| التعريف | الكتابة الصوتية |
| ترجمة | الحقل الدلالي |
| التحليل الصرفي | تاريخ التحديث |
| التحليل النحوي | ملحوظات |

الجدول ٥: أمثلة لمكونات المدخل المعجمي المختص.

كما قدمت (ISO) عددًا من وثائق المعايير الأخرى التي تهدف إلى تحسين جودة المعجم المختص واتساقه وسهولة استخدامه وتبادل ومشاركة البيانات المعجمية المختصة. فتضمنت وثيقة (ISO 26162-1:2019) مجموعة من المعايير لتصميم وإدارة قواعد بيانات المصطلحات، وإرشادات مخصصة لإنشاء وإدارة قواعد بيانات المصطلحات، تشتمل على معايير تخزين وتنظيم المصطلحات وصياغة التعريفات التي تشكل أساس المعاجم المختصة. وتركز الوثيقة في الهيكل الفني والمبادئ التنظيمية اللازمة لبناء قاعدة بيانات اصطلاحية تلبي احتياجات المستخدمين. ويمكن اعتماد عدد من معايير هذه الوثيقة لتصميم هيكل قاعدة البيانات لتخزين المصطلحات، والتعريفات، والمرادفات، والأضداد، والبيانات الوصفية الخاصة بها، مثل: المصدر، ومجال الاستخدام، واللغة، وغيرها. كما تدعم الوثيقة التوسع والتكامل مع الأدوات الأخرى، مثل: (برامج ذاكرة الترجمة، أو أنظمة إدارة المحتوى). وتؤكد أهمية تطبيق ممارسات ضمان الجودة، مثل: المراجعات الدورية لمدخلات قاعدة البيانات لضمان الاتساق، والدقة في البيانات المعجمية.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

وتتضمـن وثيقـة (ISO 22128:2008) اسـتعراضًا لأفضـل الممارسـات في بنـاء وتقديـم المنتجـات والخدمـات المتصلـة بالمداخـل المعجميـة المختصـة، مـع التركـيز في أهميـة ضمـان الجـودة والتوحيـد القيـاسي للمصطلحـات. ويمكـن أن يُفـاد مـن هـذه الوثيقـة في بنـاء إطـار عمـل لتخطيـط وإنتـاج المعاجـم المختصـة، بمـا في ذلـك تحديـد نطاقهـا، والفئـات المسـتهدفة، وصيـغ تقديمهـا (مثـل النسـخ المطبوعـة أو الرقميـة أو عـبر الإنترنـت)، وتتضمـن توصيـات الوثيقـة ضـرورة الالـتزام بمعايـير الجـودة من حيـث سـهولة الاسـتخدام، والدقة، والشـمولية. وتصميـم عـدد مـن الخدمـات المسـاندة للمعجـم، مثـل التدريـب أو أدلـة الاسـتخدام، لزيـادة مسـتوى الإفـادة الفعالـة مـن المعجـم المختـص.

ويقـدم معيـار (ISO 15188:2001) منهجًا لإدارة المشـاريع الخاصـة بتوحيد المصطلحات. يتضمـن أهـم العمليـات وأفضـل الممارسـات لتحقيـق الكفـاءة والاتسـاق في بنـاء المعجـم المختـص، ويسـاعد على تطبيـق إرشـادات إدارة المشـاريع؛ لتنظيـم عمليـة إنشـاء المعجـم، بمـا في ذلـك تحديـد الجـداول الزمنيـة، والأدوار، والمسـؤوليات للمسـاهمين في بنـاء المعجـم، مثـل: الخـبراء المختصـين، واللغويـين، والمحرريـن. كمـا يقـترح مسـارات عمـل لجمـع المصطلحـات، والتحقـق منهـا، والموافقـة على إضافتهـا للمعجـم. ويبـين أفضـل الممارسـات لإدارة مدخـلات أصحـاب المصلحـة وضمـان التعـاون بـين الفـرق المختلفـة، مثـل: الخـبراء في المجـال والمهنيـين اللغويـين والتقنيـين وغيرهـم.

وتركـز وثيقـة (ISO 860:2007) في تنسـيق بنـاء المفاهيـم والمصطلحـات في المعاجـم المختصـة، وتقديـم حلـول للاختلافـات الاصطلاحيـة وضمـان اتسـاق تعريفـات المصطلحـات ووضـوح دلالتهـا. ويمكـن الإفـادة مـن معايـير هـذه الوثيقـة في بنـاء منهجيـة لمعالجـة الاختلافـات بـين المصطلحـات أو تعـدد التعريفـات بـين التخصصـات، أو المجـالات، أو اللغـات. ويمكـن تطبيـق بعـض المعايـير لضمـان أن المصطلحـات المسـتخدمة في المعجـم المختـص متوافقـة مـع المعايـير أو المعاجـم المماثلـة في المجـال نفسـه، ويسـاعد اتبـاع إرشـادات الوثيقـة على تجنـب التكـرار أو التناقضـات في التعريفـات، وزيـادة وضـوح المداخـل المعجميـة المختصـة واتسـاقها.

وأخـيرًا، يتطلـب بنـاء معجـم إلكـتروني متخصـص اسـتخدام منهجيـات ومعايـير موحـدة لضمـان الدقـة، والاتسـاق، وسـهولة الاسـتخدام. وتسـاعد وثائـق معايـير (ISO) الـتي اسـتعرضنا أهمها في هـذا القسـم على تصميـم قاعـدة بيانـات قويـة؛ لتحليـل وتوحيـد المفاهيـم الاصطلاحيـة، وضمـان جـودة المحتـوى، وإدارة المشـاريع المعجميـة الإلكترونيـة بطريقـة منهجيـة ومنظمـة.

## نطاق المعايير وحدودها

توضـح هـذه الوثيقـة المعايـير الأساسـية لبنـاء وتقييـم المعاجـم الإلكترونيـة المختصـة، مـن خـلال تحديـد المبـادئ والخطـوات الرئيسـة الـتي يلـزم اتباعها في مشـاريع هـذا النـوع مـن المعاجـم، وتشـمل المعايـير فئـات متعـددة، مثـل: أهـداف المعجـم، وفئاتـه المسـتهدفة، إضافـة إلى نطـاق المعجـم وحـدوده، وغيرهـا مـن المعايـير الضروريـة الـتي تضمـن بنـاء معجـم مختـص متميـز، يلـبي احتياجـات مختلـف الفئـات المسـتهدفة.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

## قائمة معايير بناء المعاجم المختصة(٢):

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| **أولًا: معايير عامة لإعداد المعجم الرقمي المختص** | ١. توافر المتطلبات الأساسية لبناء وتصميم المعجم الرقمي المختص | أ. تحديد أهداف المعجم المختص ونطاقه والفئات المستفيدة. | توضيح الأهداف الرئيسة التي يسعى المعجم المختص إلى تحقيقها، مثل: تسهيل الوصول إلى مصطلحات متخصصة، أو تحسين فهم مفاهيم معينة. إضافة إلى تحديد نطاق هذا المعجم، والفئات المستهدفة، كالباحثين، أو الطلاب، أو المهنيين. | ظهور أهداف المعجم المختص ونطاقه والفئات المستهدفة | مراجعة وثائق المشروع عن طريق قائمة التحقق. |
| | | ب. تحديد احتياجات وسمات الفئات المستفيدة من المعجم المختص. | يتطلب هذا المعيار فهماً دقيقاً لاحتياجات الفئات المستهدفة، مثل: مستوى المعرفة المطلوبة، ونوعية المصطلحات التي يحتاجونها، وكيفية استخدامهم للمعجم؛ لضمان تصميم محتوى معجمي ملائم. | نسبة تغطية الاحتياجات للفئات المستفيدة عن طريق المعجم | تحليل احتياجات - استبيانات الفئات |
| | | ج. تحديد القيمة المضافة للمعجم المختص. | يشير المعيار إلى تحديد مشكلة أو فجوة محددة يسعى المعجم المختص إلى تجسيرها، مثل: نقص الموارد اللغوية المتاحة، أو صعوبة الوصول إلى المصطلحات المتخصصة في مجال معين، وتوضيح كيف يسهم المعجم المختص في حل هذه المشكلة. | نسبة وضوح الفجوة أو المشكلة التي يعالجها المعجم | مراجعة الأهداف - تحليل فجوات - مراجعة خبراء |

(٢) تم إدراج مؤشرات أداء متوسطة، لا تتطلب - غالبا - إجراءات مطولة، بحيث يمكن الإفادة منها بيسر وسهولة.

هذه الطبعة إهداء من المجمع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | د. تحديد منهج كتابة التعريفات المعجمية (المنهج الاستقرائي أو التعليمي الموجه). | تحديد منهجية واضحة لكتابة التعريفات، سواء استقرائيًا كان المنهج (يعتمد على تقديم الأمثلة أولاً ثم الوصول إلى التعريف)، أم تعليميًا موجهًا (يبدأ بالتعريف ثم يوضح الأمثلة)؛ لضمان الاتساق في تقديم المعلومات. | وجود منهج للكتابة | مراجعة وثائق - مقابلة فريق التحرير |
| | | هـ. تحديد الموضوعات التي يشتمل عليها نطاق المعجم المختص. | تحديد النطاق الدقيق للمعجم المختص والموضوعات أو المجالات التي يشملها، سواء علمية كانت الموضوعات، أم تقنية، أم أدبية؛ ليوضح للمستخدمين ما يمكنهم توقعه من محتوى المعجم المختص. | نسبة الموضوعات المحددة بوضوح | مراجعة وثائق - استبانة مستخدمين |
| | | و. تصميم وظائف المعجم المختص ومكوناته ومحتواه بناء على الأهداف واحتياجات فئات المستخدمين. | تصميم مكونات المعجم المختص -مثل: محرك البحث، والتعريفات، والأمثلة، والأدوات التفاعلية-، بناءً على الأهداف المحددة واحتياجات الفئات المستفيدة؛ بما يضمن سهولة الوصول إلى المعلومات المطلوبة. | نسبة الوظائف والمكونات المطوّرة بما يتواءم مع الأهداف واحتياجات المستفيدين. | مراجعة تقنية - اختبار قابلية الاستخدام |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ز. إعداد دليل مرجعي لمستخدم المعجم المختص. | إعداد دليل شامل يشرح للمستخدمين كيفية استخدام المعجم المختص، بما في ذلك الأهداف، والفئات المستفيدة، وطريقة جمع البيانات، ووظائف المعجم المختص المختلفة؛ مما يسهل على المستخدمين فهم كيفية استخدام المعجم بفعالية. | توفر دليل شامل وواضح | مراجعة وثائق - تجربة مستخدمين |
| **ثانيًا: معايير تجربة المستخدم** | ١. تقديم الدعم اللازم للمستخدم لمساعدته على الإفادة من المعجم المختص | أ. سهولة الوصول إلى قنوات الدعم والمساندة في المعجم المختص. | توفير قنوات الدعم والمساندة بما يسهل الوصول إليها من مختلف مكونات المعجم المختص، مثل روابط المساعدة أو الدعم الفني. | توفر قنوات الدعم في واجهة المعجم | مراجعة تقنية - اختبار عملي لواجهة المعجم المختص. |
| | | ب. وضوح بيانات المساعدة ومناسبتها لمواضع الدعم المختلفة. | أن تكون بيانات المساعدة موجهة توجيهاً دقيقاً لكل موقف، ليحصل المستخدم على الإرشادات المناسبة في الوقت والمكان الصحيحين داخل المعجم المختص. | نسبة وضوح وملاءمة بيانات المساعدة | تحليل محتوى - استبانة مستخدمين |
| | | ج. إضافة تدريب تفاعلي لطرق استعمال المعجم المختص والبحث في بياناته. | تقديم تدريب تفاعلي يساعد المستخدمين على فهم كيفية استخدام المعجم المختص بفعالية، ولا سيما طرق البحث والوصول إلى المعلومات المعجمية. | توفر وحدات التدريب التفاعلي | مراجعة تقنية - تتبع تفاعل المستخدم |

هذه الطبعة إهداء من المجمع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | ٢. تخصيص تجربة المستخدم | أ. تقديم الخدمات المخصصة للفئات المستفيدة من المعجم المختص. | تكييف المعجم المختص لتلبية احتياجات المستخدمين المختلفة، سواء مبتدئين كانوا أم خبراء، عن طريق تقديم خيارات مخصصة تناسب مستوى كل فئة. | نسبة الخدمات المخصصة في المعجم المختص، المقدمة للفئات المستفيدة. | مراجعة وظائف - مقابلات مستخدمين |
| | | ب. إمكانية إنشاء ملف تعريفي لمستخدم المعجم المختص، يحفظ اختياراته وسجل استعماله للمعجم وتفضيلاته. | تمكين المستخدمين من إنشاء ملف شخصي يحفظ تفضيلاتهم وسجل استعمالهم؛ مما يسهل عليهم الوصول إلى المحتوى والميزات المفضلة لديهم. | نسبة المستخدمين الذين تمكنوا من إنشاء ملفات شخصية نشطة | تقارير النظام - تحليل الاستخدام |
| | | ج. إمكانية إعادة تصميم وتقديم محتوى المعجم المختص وفقًا لتحليل الملف التعريفي للمستخدم. | تخصيص المحتوى بناءً على تحليل الملف التعريفي للمستخدم؛ مما يتيح تقديم تجربة مخصصة وفقًا لاحتياجاته وتفضيلاته. | نسبة نجاح (إعادة تقديم) تخصيص المحتوى بناءً على الملفات التعريفية للمستخدمين | اختبار عملي - استبيان رضا |
| | ٣. تلبية تصميم واجهة المعجم احتياجات المستخدمين المتنوعة | أ. مرونة واجهة المستخدم ودعمها لبيئات وطرق عرض الأجهزة الإلكترونية المختلفة. | أن تكون واجهة المعجم المختص مرنة وقابلة للتكيف مع مختلف الأجهزة (مثل: الهواتف الذكية، والأجهزة اللوحية)؛ مما يتيح تجربة سلسة على جميع المنصات. | نسبة توافق الواجهة مع الأجهزة | اختبار تقني متعدد الأجهزة |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ب. مناسبة التصميم لأهداف مستخدمي المعجم المختص، ومراعاته لأغراض استعمالهم. | تصميم المعجم المختص بما يتناسب مع احتياجات المستخدمين وأهدافهم، سواء تعليمية كانت، أم بحثية، أم مهنية. | نسبة رضا المستخدمين عن ملاءمة التصميم | استبيان مستخدمين - مقابلات |
| | | ج. الوضوح وسهولة الوصول إلى وظائف وبيانات المعجم المختص. | أن تكون وظائف المعجم المختص واضحة وسهلة الوصول، ليتمكن المستخدم من العثور السريع والسهل على المعلومات. | نسبة وضوح الوظائف وسهولة الوصول | اختبار قابلية الاستخدام |
| | | د. تحكم المستخدم في وظائف المعجم المختص وطريقة التنقل بين مكوناته المختلفة. | تمكين المستخدم من التحكم في وظائف المعجم المختص والتنقل بسهولة بين مختلف أقسامه، مثل البحث عن الكلمات أو التنقل بين التعريفات. | نسبة الوظائف القابلة للتحكم المباشر في المعجم المختص | اختبار عملي - مراجعة تقنية |
| | | هـ. تقديم المعلومات في القالب الإلكتروني المناسب لاحتياجات المستخدمين. | تقديم المعلومات بما يلبي تطلعات المستخدمين، مع تنظيم البيانات في قالب يسهل الوصول إليه والاستفادة منه. | نسبة رضا المستخدمين عن المعلومات المعروضة وفق احتياجاتهم | تحليل تعليقات - استبيان رضا المستخدمين |
| | | و. التعامل الأمثل مع أخطاء المستخدم المتوقعة أثناء تصفح المعجم المختص. | تعامُل النظام بذكاء مع الأخطاء المحتملة التي يرتكبها المستخدم، مع توفير إمكانية التراجع السهل عن هذه الأخطاء. | نسبة معالجة الأخطاء الشائعة للمستخدمين. | محاكاة أخطاء - مراجعة النظام |

هذه الطبعة إهداء من المجمع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ز. تمكين المستخدم من تخصيص بعض الوظائف والبيانات، وفقًا لاحتياجه وطبيعة استعماله للمعجم. | تمكين المستخدمين من تخصيص تجربتهم، عن طريق تعديل واجهة المستخدم، أو تخصيص البيانات المعروضة بما يتناسب مع احتياجاتهم. | نسبة الوظائف القابلة للتخصيص | مراجعة تقنية - اختبار عملي |
| | | ح. بساطة تصميم المعجم بطريقة يَسْهُل معها اتباعُ التعليمات المتعلقة بوظائفه ومكوناته وفهمها. | أن يكون التصميم بسيطًا وسهل الفهم، مع تعليمات واضحة تجعل من السهل على المستخدمين فهم الوظائف الأساسية للمعجم المختص. | وضوح التصميم والتعليمات المقدمة للمستخدم. | مراجعة نصوص - استبانة مستخدمين |
| | | ط. التفاعل مع المستخدم وتقديم التغذية الراجعة الفورية والمناسبة. | تفاعُل المعجم المختص مع المستخدمين فورياً، مع تقديم تغذية راجعة مناسبة حول أداء البحث، أو استخدام أي وظيفة من وظائف المعجم. | مدى توفر خاصية تقديم التغذية الراجعة للمستخدم | تتبع النظام - اختبار عملي |
| | | ي. مراعاة قدرات الذاكرة قصيرة المدى للمستخدم. | تصميم واجهة المعجم بحيث لا تقدم أكثر من خمسة عناصر على الشاشة في الوقت نفسه؛ مما يسهل على المستخدم التركيز على المعلومات المعروضة. | نسبة تقليل الحمل المعرفي | مراجعة تصميم - اختبار عملي |
| | | ك. مراعاة التعددية الثقافية في تصميم واجهة المستخدم. | أن تكون واجهة المعجم المختص متعددة الثقافات، لتناسب المستخدمين ذوي الخلفيات الثقافية واللغوية المختلفة. | نسبة توافق التصميم مع ثقافات متعددة | مراجعة خبراء - تحليل تصميم |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | ٤. دعم عمليات الوصول والتنقل بين مكونات المعجم المختص، ومشاركة بياناته | أ. الامتثال لمعايير الوصول لمحتوى المواقع الإلكترونية، ولا سيما من قبل ذوي الاحتياجات الخاصة. | أن يلتزم المعجم بمعايير الوصول، مثل WCAG، لضمان إمكانية الوصول لذوي الاحتياجات الخاصة، بما في ذلك قارئات الشاشة. | نسبة توافق المنصة مع معايير WCAG | تدقيق نفاذية باستخدام أدوات فحص + اختبار مع مستخدمين ذوي إعاقة |
| | | ب. سهولة التنقل بين مكونات المعجم المختص المختلفة. | سلاسة التنقل بين أقسام المعجم المختص المختلفة؛ مما يساعد المستخدم على الوصول إلى المعلومات المطلوبة بسهولة. | متوسط عدد النقرات للوصول إلى المعلومة | اختبار عملي لسيناريوهات التنقل |
| | | ج. تصميم محرك بحث بمواصفات تقنية متقدمة. | تصميم محرك البحث بطريقة تتيح الوصول إلى المعلومات المعجمية بأسرع وقت وأقل جهد ممكن، باستخدام تقنيات بحث متقدمة. | متوسط زمن الاسترجاع (ثوانٍ) | اختبارات أداء للنظام |
| | | د. تقديم الدعم الملائم للمستخدم لاختيار الطريقة الأنسب في الوصول إلى البيانات. | يُقدّم المعجم المختص إرشادات أو توصيات حول أفضل طريقة لاستخدام أدوات البحث أو اختيار البيانات المناسبة. | مدى توفر الإرشادات والدعم المناسب للمستخدم | استبيانات وتجارب استخدام عملية |
| | | هـ. التمييز بين روابط التنقل الداخلية والخارجية في المعجم المختص. | التمييز الواضح بين الروابط الداخلية (التي تقود إلى أقسام أخرى داخل المعجم المختص) والخارجية (التي تقود إلى مصادر خارجية). | نسبة وضوح الروابط الداخلية والخارجية | مراجعة واجهة المعجم وفحص الروابط |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | و. استخدام الروابط بطرق مناسبة ومساعدة على تنقل منطقي وسلس بين مكونات المعجم المختص. | استخدام الروابط بطريقة تسهل التنقل المنطقي والسلس بين أجزاء المعجم المختص المختلفة؛ مما يعزز تجربة المستخدم. | معدل رضا المستخدمين عن منطقية الروابط | استبيانات + تتبع جلسات الاستخدام |
| | | ز. استعمال الروابط التشعبية بكفاءة في داخل تعريف المدخل المعجمي وعلاقاته الدلالية. | استخدام الروابط التشعبية بفعالية داخل التعريفات لربط المصطلحات والمعاني ذات الصلة؛ مما يسهل على المستخدم التنقل بين المفاهيم الدلالية المتصلة. | نسبة المداخل التي تستعمل روابط تشعبية صحيحة | مراجعة عينة من المداخل للتحقق من الروابط |
| | | ح. دعم المعجم المختص الوصول للموقع والاستخدام من خلال لوحة المفاتيح. | أن يدعم المعجم أدوات تسهيل الوصول، مثل إمكانية التحكم في الموقع عبر لوحة المفاتيح؛ لتسهيل التصفح على جميع الفئات. | نسبة الوظائف القابلة للوصول والاستخدام عبر لوحة المفاتيح | اختبار عملي للتنقل باستخدام لوحة المفاتيح فقط |
| | | ط. إتاحة الوصول إلى النسخ السابقة من المعجم. | إتاحة إمكانية الوصول إلى النسخ السابقة من المصطلحات أو المداخل المعجمية؛ مما يسمح للمستخدمين بمراجعة التعديلات السابقة. | مدى توفر إمكانية الوصول والاطلاع على النسخ السابقة من المعجم بسهولة. | مراجعة المحفوظ الرقمي واختبار الوصول |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ي. إتاحة قنوات أو واجهة برمجية (API) لمشاركة ودمج البيانات المعجمية للمستخدمين والمطورين. | تتوافر واجهة برمجية تطبيقية (API) تتيح للمطورين والمستخدمين دمج البيانات المعجمية، أو الإفادة منها في تطبيقات أخرى؛ مما يسهل تبادل البيانات. | نسبة نجاح استدعاءات الـAPI | اختبار API في بيئات مختلفة مع مطورين |
| ثالثًا: معايير المدخل المعجمي المختص | ١.تعريف المدخل المعجمي | أ. صياغة المصطلح بأسلوب بسيط ومناسب للفئات المستفيدة. | كتابة المصطلح في شكل اسم مفرد أو جملة اسمية؛ لتسهيل الفهم، مع تجنب البدء بفعل أو حرف قدر الإمكان. | نسبة التعريفات السليمة والواضحة | مراجعة لغوية - اختبار مستخدمين |
| | | ب. ارتباط المدخل المعجمي المختص بمفهوم واحد في المجال المختص مع وضوح دلالته. | أن يعبر كل مدخل معجمي عن مفهوم واحد في المجال المختص، مع مراعاة معالجة المفاهيم متعددة كمداخل مستقلة. | نسبة التعريفات ذات المعنى الواحد في المجال المختص | مراجعة خبراء - مقارنة مصادر |
| | | ج. ربط مفهوم المصطلح بالمفاهيم والمداخل المعجمية المختصة ذات العلاقة (علاقة هرمية أو ترابطية). | ربط المصطلحات بمفاهيم أخرى ذات علاقة، سواء هرمية كانت العلاقة (مثل التصنيف)، أم ترابطية (مثل الترادف). | نسبة الروابط بين المصطلحات والمداخل | مراجعة خبراء - تحليل دلالي |
| | | د. تبني الترجمات العلمية المعتمدة من المجامع والجهات المختصة عند نقل المصطلحات من اللغات الأخرى في كتابة المداخل المعجمية المختصة. | أن يعتمد المعجم في نقل المصطلح الأجنبي على الترجمة التي اعتمدتها الجهات اللغوية والمجامع العلمية ذات العلاقة. | نسبة التعريفات المترجمة المعتمدة | مراجعة محتوى - تدقيق داخلي |

هذه الطبعة إهداء من المجمع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | هـ. استخراج الأمثلة المناسبة للمداخل المعجمية المختصة من النصوص اللغوية المختصة التي تعد المصدر الأهم لأمثلة المداخل المعجمية. | أن تستمد أغلب المصطلحات أمثلتها من سياقات استعمالها في النصوص اللغوية المتخصصة التي تعد المصدر الرئيسي لفهم وشرح المداخل المعجمية المختصة. | نسبة الأمثلة ذات المرجعية الصحيحة | مراجعة محتوى - استبيان مستخدمين |
| | | و. استعمال أساليب متعددة في التعريف بحسب ما يناسب المدخل المعجمي المختص. | استخدام طرق متعددة للتعريف بالمصطلح، ومن أهمها: التعريف بالمثال، أو المرادف، أو الضد، أو بالتضمين عن طريق تقديم شرح عام للمصطلح ثم ذكر أهم السمات المميزة له. | نسبة المصطلحات المستخدمة لأساليب متنوعة في التعريف. | مراجعة خبراء - تدقيق أكاديمي |
| | | ز. الإيجاز ووضوح اللغة التي يكتب التعريف بها. | يُفضل تضمين التعريف أهم سمات المصطلح الأساسية بلغة موجزة سهلة الفهم؛ لشرح المعنى شرحاً واضحاً ودقيقاً لغير المختص. | نسبة التعريفات الموجزة | مراجعة ببليوغرافية - مقارنة مصادر |
| | | ح. التأكد من صحة مفهوم المصطلح ودقته في الدلالة على المعنى. | يفضّل التحقق من استخدام المصطلحات ذات الدلالات الواضحة، والمعاني المباشرة، والبعد عن استخدام المصطلحات ثائية الدلالة. | نسبة المداخل الدقيقة | مراجعة محتوى - تحليل خبراء |
| | | ط. كتابة التعريف بصيغة إيجابية. | يُفضل صياغة التعريفات بأسلوب الإثبات (الإيجابي)، وتجنب أسلوب النفي. | نسبة المداخل المكتوبة بصيغة إيجابية | مراجعة لغوية - اختبار عملي |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ى. تجنب استخدام المصطلح نفسه في التعريف به. (دوران التعريف). | أن يُكتب التعريف دون أن يُكرر المصطلح نفسه في الشرح؛ أي ينبغي تجنب استخدام المصطلح نفسه في تعريفه. | نسبة المداخل الدائرة على نفسها (النسبة بالسالب) أي كلما كانت اقل كلما كان افضل | مراجعة لغوية - اختبار عملي |
| | | ك. اتباع منهج ثابت وواضح في صياغة التعاريف في المعجم المختص. | أن يكون هناك اتساق في طريقة صياغة التعريفات، بأن يتبع كل تعريف منهجية واضحة وثابتة. | وجود المنهج | مراجعة خبراء - تحليل تصنيفي |
| | ٢. استيفاء المكونات الأساسية في شرح المدخل المعجمي | أ. موافقة مكونات المدخل المعجمي فئات البيانات في المعايير العالمية للمعاجم المختصة. | أن تتوافق مكونات المدخل المعجمي مع المعايير العالمية الخاصة بالمعاجم المختصة. | نسبة المداخل المتوافقة | مراجعة لغوية - تدقيق داخلي |
| | | ب. اشتمال المدخل المعجمي المختص على العناصر الرئيسة | أن يحتوي المدخل المعجمي المختص على العناصر الرئيسة، مثل: عنوان المدخل المعجمي، والتعريف، والأمثلة السياقية. | نسبة المداخل المكتملة من حيث العناصر الرئيسة. | مراجعة لغوية - مقارنة مصادر |
| | | ج. اشتمال المدخل المعجمي على ما يناسب من الفئات الإضافية، وفقًا لمعيار آيزو ١٢٦٢٠. | يمكن أن يشتمل المدخل على مكونات إضافية تتضمن أكثر من ٢٠ نوعًا من البيانات، كما ذكرت بالتفصيل في معيار آيزو ١٢٦٢٠. | نسبة المداخل التي تعرض الفئات الإضافية | مراجعة صرفية - تحليل لغوي |
| | | د. الربط المنطقي بين مكونات المدخل المعجمي المختص، وتخصيص إعدادات ما يظهر من المكونات للمستخدم. | أن تكون هناك علاقة منطقية بين مكونات المدخل المعجمي، مع إمكانية تخصيص طريقة عرض هذه المكونات وفقًا لاحتياجات المستخدم. | نسبة الترابط المنطقي في المداخل المعروضة. | مراجعة صرفية - تحليل لغوي |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | هـ. إتاحة أدوات الإكمال التلقائي لبيانات مكونات المداخل المعجمية. | إتاحة أدوات الإكمال التلقائي للمحررين؛ لتسهيل إدخال البيانات وتصحيحها بسرعة ودقة. | توفر خاصية أدوات الإكمال التلقائي وفاعليتها. | مراجعة تقنية - اختبار عملي |
| | | و. اشتمال المدخل المعجمي المختص على معلومات معرفية إثرائية عامة خارج الشرح اللغوي للمدخل (الإثراء الثقافي للمدخل المعجمي). | من الجيد إضافة معلومات معرفية خارج الشرح اللغوي، مثل الحقائق الثقافية أو التاريخية المرتبطة بالمدخل المعجمي المختص. | نسبة المداخل التي تعرض معلومات إثرائية | مراجعة لغوية - تحليل دلالي |
| | | ز. تنظيم المعلومات في المدخل المعجمي بطرق متسقة تساعد على سهولة استعماله. | ينبغي تنظيم المعلومات داخل المدخل المعجمي تنظيماً متسقاً، ليسهل على المستخدمين العثور على البيانات المطلوبة. | وجود تنظيم للمعلومات في المعجم | مراجعة- تحليل نصوص |
| | | ح. المقارنات المرجعية لمكونات المعجم المختص مع المصادر المناظرة محليًا وعالميًا. | ينبغي مقارنة مكونات المعجم المختص مع مصادر أخرى محلية وعالمية؛ لضمان دقة المعلومات وشمول التناول. | نسبة المداخل التي تحتوي مقارنات | مراجعة ثنائية اللغة - مقارنة مصادر |
| | | ط. دعم المعجم المختص للتعدد اللغوي. | أن يدعم المعجم المختص التعدد اللغوي؛ مما يتيح ترجمة المصطلحات إلى لغات عدة. | وجود ترجمة | مراجعة تقنية - اختبار عملي |
| | | ي. تقديم التحليل اللغوي للوحدة المعجمية المختصة (الصرفي، النحوي). | أن يشمل المعجم المختص تحليلًا لغويًّا دقيقًا للوحدات المعجمية، بما في ذلك التحليل الصرفي والنحوي. | توفر خاصية التحليل اللغوي | مراجعة خواص المنصة - اختبار عملي |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | ٣. جـــــودة البيانـــــات الوصفيـــــة وشـمولها | أ. كتابـة البيانـــات الوصفيـــة وفـقًا للنـموذج المعتمـد لتمثيـل البيـانات المعجمية المختصة. | كتابة البيانات الوصفية بما يتوافق مع النماذج المعتمدة لتمثيل البيانات المعجمية المختصة. | نسبـة المـداخل الخالـيـــة مـن الأخطاء | مراجعة لغوية - تدقيق داخلي |
| | | ب. اشتمال البيانات الوصفية على معلومات حول المصدر وتاريخ الإضافة. | أن تتضمن البيانات الوصفية المختصة معلومات حول مصدر المصطلح وتاريخ إضافته. | نسبة المداخل الشاملة للمصدر والتاريخ | مراجعة خبراء - تحليل محتوى |
| | | ج. تضمين بيانات المؤلف أو مدخل البيانات. | أن يحتوي المعجم المختص على معلومات حول مؤلفي البيانات أو مدخلي البيانات. | نسبة المداخل الشاملة للمؤلف أو مدخل البيانات | مراجعة خبراء - تحليل محتوى |
| | | د. إمكانية تتبع تحديثات المدخل المعجمي المختص. | أن يتضمن المعجم المختص إمكانية تتبع التحديثات التي أجريت على المدخل المعجمي، سواء ما يتعلق بتاريخ تحديثها أم نوع البيانات المحدثة. | وجود متابعة تاريخية لتحديثات المدخل | مراجعة سجلات - مقارنة إصدارات |
| **رابعًا: معايير المحتوى المعجمي** | ١. ضمان جودة محتوى البيانات المعجمية المختصة | أ. مناسبة المحتوى لأهداف المعجم المختص والفئات المستفيدة. | أن يتوافق محتوى المعجم المختص مع الأهداف المحددة ومع احتياجات الفئات المستفيدة، كالطلاب أو الباحثين في المجال المختص. | نسبة توافق المحتوى مع أهداف المحددة واحتياجات الفئات المستفيدة. | تحليل محتوى |
| | | ب. شمول المداخل المعجمية المختصة ضمن نطاق المعجم من حيث المفاهيم والحجم. | أن يتناول المعجم جميع المفاهيم ذات الصلة وبحجم مناسب يضمن شمولية المداخل المعجمية ضمن نطاق المعجم المحدد. | نسبة تغطية المفاهيم الأساسية المعتمدة | مقارنة قائمة المفاهيم الأساسية (مرجعية) مع محتوى المداخل ورصد الفجوات |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ج. مناسبة مستوى التفاصيل وتعقيد المعلومات المتصلة بالمدخل المعجمي المختص لأهداف المعجم. | أن تكون المعلومات المقدمة متوازنة بين التفصيل والبساطة، لتتناسب مع مستوى الفهم المطلوب للفئات المستفيدة. | نسبة التوازن بين التفصيل والبساطة بما يتناسب مع الأهداف | تطبيق قائمة تحقق لمستويات التفصيل (مبتدئ/ متوسط/ متقدم) بحسب بطاقة التصميم |
| | | د.اعتماد المعجم على مدونات لغوية مختصة عالية الجودة وملائمة لأهداف المعجم. | أن يعتمد المعجم على مدونات لغوية مختصة تحتوي على بيانات موثوقة وذات جودة عالية تتلاءم مع أهداف المعجم. | نسبة المدونات اللغوية المختصة عالية الجودة التي يعتمد عليها المعجم. | مراجعة الخبراء – تحليل المحتوى. |
| | | هـ. مناسبة لغة المدونة المختصة للمعجم من حيث النوع والمستوى اللغوي لأهداف المعجم والفئات المستفيدة. | أن تكون لغة المدونة المختصة متوافقة مع مستوى المستخدمين -سواء مبتدئين كانوا أم متخصصين-، وأن تلبي احتياجات الفئة المستهدفة. | نسبة المدونات المتوافقة مع مستوى المستخدمين. | مراجعة الخبراء تحليل المحتوى. |
| | | و. جودة الأمثلة المستعملة في تعريف المدخل المعجمي ودقتها في تمثيل استعمال المعنى. | أن تكون الأمثلة المقدمة في المعجم دقيقة، وتعكس الاستخدام الفعلي للمصطلحات في سياقاتها الحقيقية. | نسبة الأمثلة الصحيحة والدالّة على المعنى | تحكيم لغوي ودلالي للأمثلة وفق سلم ترميز (صحيح/جزئي/ غير صحيح) |
| | | ز. الربط بمصادر خارجية إثرائية متصلة بالمداخل المعجمية المختصة. | أن يقدم المعجم روابط لمصادر خارجية موثوقة تعزز الفهم وتقدم معلومات إضافية للمستخدمين، مثل مواقع الموسوعات والمصادر المعرفية. | نسبة المداخل المرتبطة بمصدر إثرائي واحد على الأقل | فحص الروابط (صلاحية / ملاءمة/دقة) وتتبعها آليًا + تحقق بشري لعينة |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ح. تمكين المستخدم من تقييم المعلومات المعجمية المختصة والتعليق عليها. | أن يتمكن المستخدمون من تقييم المحتوى المعجمي والتعليق عليه لتقديم ملحوظات أو اقتراحات لتحسين المحتوى. | وجود خاصية تقييم المستخدم للمعلومات والتعليق عليها. | مراجعة تقنية- اختبار عملي |
| | | ط. التحديث المستمر لمحتوى المعجم المختص ومكونات المداخل المعجمية. | ينبغي تحديث المعجم دورياً لمواكبة التغيرات في المصطلحات والمعاني والبيانات الجديدة في المجال المختص. | نسبة المداخل المُحدَّثة خلال آخر ١٢ شهرًا | مراجعة سجل الإصدارات/ التغييرات (changelog) وفترات إعادة التحرير الدورية |
| | | ي. وضوح اللغة وسلامتها من الأخطاء. | أن تكون اللغة المستخدمة في التعريفات واضحة وخالية من الأخطاء. | نسبة المداخل الخالية من الأخطاء اللغوية والإملائية | تدقيق آلي أولي + تدقيق بشري ثانٍ وفق دليل التدقيق اللغوي |
| | | ك. تجنُّب تضمين الاختصارات اللغوية غير الضرورية. | اجتناب ذكر الاختصارات التي قد تربك المستخدم وتسبب الغموض في فهم المصطلح. | نسبة الاختصارات اللغوية في المداخل. | مراجعة لغوية (كلما قلت النسبة كان أفضل) |
| | | ل. اتساق العلاقة ومناسبتها بين العناوين المستعملة والمحتوى الخاص بها. | أن تكون العناوين المستخدمة متسقة مع المحتوى المرتبط بها؛ مما يسهل على المستخدمين فهم المعلومات. | نسبة العناوين المتسقة دلاليًا مع المحتوى | تحليل المحتوى |
| | | م. الاستعمال الفعّال للوسائط المتعددة لتوضيح المعاني والعلاقات المعجمية. | ينبغي استخدام الوسائط المتعددة، مثل: الصور والفيديوهات؛ لتوضيح المعاني والعلاقات بين المصطلحات بطريقة تفاعلية. | نسبة المداخل التي توظّف وسائط متعددة مناسبة عند الحاجة | تقييم جودة الوسائط (وضوح/ ترخيص/ ملاءمة) وربطها بالمعنى المستهدف |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ن. توظيف المعلومات الإحصائية لتصنيف المداخل المعجمية بطرق مرئية (التصنيف بحسب الشيوع - تقنية سحابة الكلمات - دعم التصنيف الدلالي). | يمكن استخدام الإحصاءات مباشرة في تصنيف المداخل المعجمية بحسب درجة شيوعها أو عرضها في سحب الكلمات والرسوم البيانية. أما التصنيف الدلالي، فهو يقوم أساسًا على العلاقات المفهومية بين الوحدات المعجمية (مثل: علاقة النوع والفئة، أو الجزء والكل)، ويمكن للإحصاءات أن تُستعمل هنا كأداة مساعدة في كشف أنماط التشارك السياقي أو الترابط في الاستعمال؛ مما يعزز بناء الحقول الدلالية بطريقة أكثر موضوعية ودقة. | نسبة المداخل المزوّدة بتصنيفات / مرئيات إحصائية صحيحة | مراجعة الخبراء- تحليل المحتوى. |
| | | س. مشاركة الخبراء في المجال المختص الذين يمتلكون الخبرات والمعارف والمهارات الخاصة بمحتوى المعجم. | أن يمتلك مطورو المعجم الخبرة والمعرفة الكافية في صناعة المعاجم المختصة؛ لضمان الجودة والدقة في المحتوى المختص. | نسبة المداخل التي خضعت لتحكيم خبير متخصص | سجل تحكيم الخبراء (اسم الخبير/ تخصصه/تاريخ المراجعة) وتتبع إعادة التحرير بناءً على الملاحظات |
| | ٢. تحديد نطاق محتوى المعجم وشموله | أ. تحديد المجال الدقيق لمحتوى المعجم المختص والفئات المستفيدة. | ينبغي تحديد المجال الدقيق الذي يشمله المعجم، والفئات المستفيدة منه؛ لضمان توافق المحتوى مع احتياجات المستخدمين. | نسبة المحتوى الموسوم بالمجال والفئات | مراجعة خبراء - تحليل محتوى |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ب. شمول المعجم لجميع المداخل المعجمية المرتبطة بالمجال المختص. | أن يشمل المعجم جميع المصطلحات والمفاهيم المتعلقة بالمجال المختص؛ لضمان اكتمال المحتوى، ولتوفير مرجع شامل. | نسبة المداخل المكتملة | مراجعة خبراء - تحليل محتوى |
| | | ج. وجود قائمة معايير مبنية على المحتوى المختص لإضافة أو حذف المداخل المعجمية التي يتضمنها المعجم. | وجود معايير واضحة لإضافة أو حذف المداخل، بناءً على أهميتها وارتباطها بالمجال المختص. | توفر قائمة المعايير | مراجعة خبراء - تحليل محتوى |
| | ٣. ضبط جودة المحتوى وتحديثه | أ. مراجعة اللغويين والخبراء لتعريفات المداخل المعجمية وتدقيقها. | أن يدقق التعريفات ويراجعها خبراء لغويون؛ لضمان دقتها وصحتها قبل إضافتها إلى المعجم. | نسبة المداخل التي تمت مراجعتها وتدقيقها من قبل خبراء | فحص سجل المراجعات الموقّعة من اللغويين والخبراء |
| | | ب. وجود آلية لتلقي تعليقات المستخدمين وتصحيحاتهم لمحتوى المعجم المختص. | وجود آلية تتيح للمستخدمين تقديم التعليقات أو التصحيحات على المحتوى المعجمي؛ لضمان تحسينه باستمرار. | وجود خاصية التعليقات للمستخدمين. | مراجعة تقنية |
| | | ج. التحديث المستمر لمحتوى المعجم ليمثل أي تغييرات تطرأ على المعاني المعجمية وفقًا للاستعمالات اللغوية والتداولية المستجدة. | ينبغي تحديث المعجم بانتظام ليعكس التغييرات في استخدام المصطلحات والمعاني المعجمية وفقًا للاستعمالات الحديثة. | نسبة المداخل التي تم تحديثها خلال آخر ١٢ شهرًا | مراجعة سجل التعديلات والإصدارات |

هذه الطبعة إهداء من المجمع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | د. وجود دليل مرجعي لمحرري المعجم؛ لضمان اتساق كتابة البيانات المعجمية في كل المداخل التي يتضمنها المعجم. | أن يكون هناك دليل مرجعي يوضح المعايير والأساليب المتبعة في تحرير المداخل؛ لضمان الاتساق والدقة في جميع التعريفات. | توفر دليل مرجعي رسمي لمحرري المعجم | التحقق من وجود الدليل ومطابقته مع ممارسات التحرير |
| **خامسًا: نموذج تمثيل البيانات المعجمية المختصة** | ١. مراعاة جودة تصميم وبناء نموذج البيانات المعجمية | أ. موافقة المعايير العالمية في نمذجة البيانات المعجمية المختصة. | أن يتبع المعجم المعايير العالمية المعتمدة في نمذجة البيانات (مثل: قواعد البيانات المبنية على العلاقات، أو المبنية على العلاقات الخطية أو الهرمية: relational database, graph database, or ontology).<br>ومن أبرز هذه المعايير: دبلن كور (Dublin Core)، والإطار المرجعي لتوسيم البيانات المعجمية (Lexical Markup Framework – LMF). | نسبة عناصر النموذج المتوافقة مع معايير دولية مثل (TEI, LMF...) | تدقيق فني لمطابقة النموذج مع معايير دولية معتمدة |
| | | ب. توافق نموذج المعجم المختص مع أهدافه مستخدميه واحتياجاتهم. | أن يكون تصميم نموذج المعجم متلائمًا مع الأهداف المحددة للمعجم، وملبيًا لاحتياجات المستخدمين. | نسبة رضا المستخدمين والخبراء عن ملاءمة النموذج | استبيانات + مقابلات مع مستخدمين وخبراء |
| | | ج. وضوح مكونات نموذج البيانات المعجمية والعلاقات والوظائف المرتبطة بها. | وضوح مكونات نموذج البيانات المعجمية، مع تحديد العلاقات بين المصطلحات والوظائف المختلفة؛ مما يسهل على المستخدمين فهم بنية المعجم. | نسبة المكونات والعلاقات الموثقة بوضوح | مراجعة وثائق التصميم والنموذج المفاهيمي |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | د. تمثيل نموذج المعجم المختص لكافة البيانات المعجمية التي يتضمنها المدخل المعجمي. | يحتوي النموذج على كل البيانات المهمة التي تحتاج إليها المداخل المعجمية، مثل: التعريفات، والأمثلة، والتحليل اللغوي. | نسبة المداخل التي يغطيها النموذج بشكل كامل | مطابقة محتوى المداخل مع عناصر النموذج |
| | | هـ. دعم إمكانية مشاركة المعلومات المعجمية بطرق متعددة. | أن يتيح المعجم للمستخدمين إمكانية مشاركة البيانات المعجمية بطرق مختلفة، سواء عبر وسائل التواصل الاجتماعي، أم البريد الإلكتروني، أم التنزيل المباشر. | توفر طرق التصدير المتاحة | مراجعة تقنية _ اختبار مشاركة البيانات |
| | | و. سهولة إعادة استعمال بيانات المعجم المختص في نماذج مشابهة. | أن يكون تصميم البيانات مرنًا بما يسمح بإعادة استخدامها في نماذج أو معاجم أخرى؛ مما يعزز من قابلية البيانات للتكامل والاستخدام المتعدد. | نسبة نجاح إعادة الاستعمال في نماذج مختلفة | اختبار إعادة استعمال البيانات في نماذج مختلفة |
| | | ز. إمكانية الإضافة والتعديل لمكونات نموذج تمثيل بيانات المعجم المختص. | السماح بإمكانية التعديل والإضافة، ليتمكن المحررون من تحديث البيانات بسهولة، مع إضافة مصطلحات جديدة أو تعديل القديمة. | توفر خاصية التعديل أو الإضافة للبيانات | مراجعة تقنية - اختبار لإضافة وتعديل المكونات |
| | | ح. التنظيم المنطقي لمكونات المدخل المعجمي بما يناسب سياق استعمال المعجم. | ينبغي تنظيم مكونات المدخل المعجمي تنظيماً منطقياً، ليكون الوصول إلى المعلومات سلسًا ويتناسب مع استخدامات المعجم. | تنظيم مكونات المدخل بطريقة مناسبة | مراجعة هيكل المداخل باستخدام قائمة تحقق تنظيمية |

هذه الطبعة إهداء من المجمع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ط. مناسبة عرض نموذج تمثيل البيانات المعجمية على مختلف الأجهزة الإلكترونية. | تصميم عرض البيانات ليتكيف مع مختلف الأجهزة الإلكترونية، من الحواسيب إلى الهواتف الذكية والأجهزة اللوحية؛ لضمان سهولة التصفح. | نسبة نجاح العرض عبر مختلف الأجهزة | اختبار العرض على البيئات مختلفة |
| | | ي. دعم النموذج للتصميم المناسب لمظهر نموذج المعجم المختص ومكوناته المتعددة. | أن يدعم النموذج تصميمًا مناسبًا لمظهر واجهة المعجم، بما يتلاءم مع وظائفه المختلفة، ويعزز تجربة المستخدم. | نسبة توافق المظهر مع معايير التصميم المعتمدة | مراجعة التصميم البصري للنموذج وفق معايير تجربة المستخدم |
| **سادسًا: معايير البحث في البيانات المعجمية وعرضها** | ١. تصميم أداة البحث في المعجم الإلكتروني لتدعم تنوع خيارات المستخدم | أ. وضوح محرك البحث في الموقع أثناء التنقل بين مكونات المعجم المختص وصفحاته المختلفة. | أن يكون محرك البحث بارزًا ويسهل الوصول إليه من أي صفحة، ليستطيع المستخدمون استخدامه بسهولة أثناء التنقل بين مكونات المعجم. | نسبة نجاح التنقل السلس بين المكونات | اختبارات عملية – تحليل سجلات البحث. |
| | | ب. توظيف سمات البحث المتقدم في المعجم المختص عامةً، وفي محتوى المدخل المعجمي خاصةً. | أن يُقدّم المعجم خيارات بحث متقدمة تتيح للمستخدمين البحث الدقيق داخل المصطلحات والمعاني، مثل البحث بالمجال أو النوع الصرفي. | عدد خيارات البحث المتقدم المتاحة | مراجعة وظائف البحث وقائمة السمات |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ج. كفاءة وظائف البحث (الوصول إلى المعلومة المعجمية في أسرع وقت باستعمال أنسب التقنيات البحثية). | أن يكون محرك البحث سريعًا وفعالًا، بما يُمكِّن من العرض السريع للنتائج باستخدام تقنيات بحث متقدمة. | متوسط زمن الاسترجاع (بالثواني) | اختبارات أداء آلية لقياس سرعة البحث |
| | | د. تعدد الاختيارات والسمات التي يمكن استعمالها لتخصيص البحث وتيسير الوصول إلى النتائج المستهدفة. | أن يتيح البحث خيارات لتخصيص الاستعلامات، مثل: البحث بالكلمات المفتاحية، أو الفئات الصرفية، أو الموضوعات. | وجود خاصية تنوع خيارات التخصيص في البحث. | مراجعة تقنية - اختبار عملي. |
| | | هـ. دعم البحث عن الحروف والرموز والكلمات المفردة والتراكيب. | أن يدعم محرك البحث إمكانية البحث باستخدام الحروف، أو الرموز، أو الكلمات المفردة، أو التراكيب اللغوية. | وجود خاصية البحث باستخدام الحروف أو الرموز أو الكلمات المفردة أو التراكيب. | مراجعة تقنية- اختبار عملي. |
| | | و. إضافة ميزة الإكمال التلقائي للمدخلات المعجمية. | أن يقترح محرك البحث تلقائيًا الكلمات أو المصطلحات بناءً على الحروف التي يكتبها المستخدم؛ مما يسهل الوصول إلى المصطلحات المطلوبة. | وجود خاصية الإكمال التلقائي | مراجعة تقنية -اختبار عملي على محرك البحث |
| | | ز. تقديم مقترحات مقاربة للمدخات الخاطئة إملائيًا أو دلاليًا. | أن يقدم محرك البحث مقترحات بديلة إملائيًا أو دلاليًا وفقًا لأقرب الاحتمالات. | توفر خاصية التصحيح الإملائي | المراجعة التقنية - الاختبار العملي |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ح. تمكين التصحيح الإملائي الآلي للمداخل المعجمية، وإظهار البدائل المحتملة. | ينبغي دعم التصحيح التلقائي للأخطاء الإملائية، مع تقديم بدائل محتملة للكلمات التي قد تكون مكتوبة كتابة خاطئة. | توفر خاصية التصحيح التلقائي الإملائي. | مراجعة تقنية -اختبار عملي لتصحيح الأخطاء الإملائية |
| | | ط. تمكين المستخدم من حفظ سجل البحث، وإعادة استعماله، وإضافته إلى المفضلة. | أن يتمكن المستخدم من حفظ استعلامات البحث السابقة وإعادة استخدامها لاحقًا، مع إمكانية إضافة نتائج البحث إلى قائمة المفضلة. | وجود خاصية حفظ سجلات البحث للمستخدم | مراجعة تقنية - اختبار عملي. |
| | | ي. تمكين البحث الصوتي وتحويل الصوت إلى نص، وتقديم التصحيح المناسب في حالة الخطأ في التعرف الآلي على الصوت. | دعم البحث الصوتي عن طريق تقنية تحويل الصوت إلى نص، مع تصحيح تلقائي للأخطاء الناتجة عن أخطاء التعرف الصوتي. | وجود خاصية البحث الصوتي، مع التصحيح التلقائي لأخطاء التعرف. | مراجعة تقنية - اختبار عملي - تجارب إدخال صوتي ومقارنة النتائج |
| | | ك. إظهار النوع الصرفي للمدخل المعجمي في قوائم الإكمال التلقائي لمحرك البحث. | أن يُظهر محرك البحث النوع الصرفي، مثل: (فعل، اسم) للكلمات المقترحة في الإكمال التلقائي؛ لتسهيل اختيار المراد. | نسبة المداخل التي يظهر نوعها الصرفي | مراجعة عينة من نتائج البحث - اختبار عملي لمحرك البحث |
| | | ل. تمكين التعرف الآلي على الصور والنصوص المكتوبة. | أن يدعم المحرك تعرف النصوص المكتوبة في الصور؛ مما يتيح للمستخدم البحث عن المصطلحات المضمنة في الصور. | توفر خاصية تعرف محرك البحث على النصوص المكتوبة في الصور. | مراجعة تقنية - اختبار عملي. |

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | م. تمكين البحث بالمعنى أو العلاقات الدلالية المختلفة للمدخل المعجمي. | أن يدعم البحثُ القدرةَ على الاستفسار عن الكلمات، بناءً على معناها أو على العلاقات الدلالية المختلفة، مثل: الترادف أو التضاد. | توفر الخاصية في محرك البحث | مراجعة تقنية - اختبار بحث دلالي بعلاقات (ترادف / تضاد...) |
| | | ن. تمكين البحث بالقافية والسوابق واللواحق واللواصق الصرفية. | أن يتيح محرك البحث إمكانية إيجاد الكلمات التي تتطابق في القافية أو تحتوي على سوابق أو لواحق معينة أو لواصق صرفية. مثل البحث بالوزن الصرفي أو بالقافية أو الزيادات الصرفية. | توفر الخاصية في محرك البحث | مراجعة تقنية - إدخال استعلامات صرفية واختبار النتائج |
| | | س. تمكين البحث بنمط الكلمة أو التركيب. | أن يدعم المحركُ البحثَ باستخدام أنماط محددة من الكلمات أو التراكيب اللغوية، مثل البحث عن الجمل الفعلية أو الاسمية. | نسبة نجاح البحث بالنمط | اختبار أنماط كلمات وجمل مختلفة |
| | | ع. تمكين البحث بالأوزان الصرفية. | أن يتيح المحركُ البحثَ عن الكلمات بناءً على الأوزان الصرفية (مثل «فَعَلَ»، «مَفعول»). | توفر الخاصية في محرك البحث | مراجعة تقنية - اختبار إدخال أوزان صرفية ومطابقة النتائج |
| | | ف. تقديم مقترحات للمداخل المعجمية المشابهة أو المتصلة بعلاقات صرفية أو دلالية. | أن يقدم محرك البحث توصيات بكلمات مشابهة أو متصلة بالمدخلات المعجمية بناءً على العلاقات الصرفية أو الدلالية. | توفر خاصية تقديم المقترحات | مراجعة تقنية - مراجعة عينة من المقترحات تلقائيًا وبشريًا |

هذه الطبعة إهداء من المجمع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ص. استخراج جميع المداخل المعجمية المتصلة، والتمييز بينها دون فتح الصفحات الخاصة بها. | أن يعرض محرك البحث جميع المداخل المعجمية المتصلة في قائمة منسقة، دون الحاجة لفتح كل مدخل على حدة. | نسبة المداخل المستخرجة بدقة | اختبار آلي لخاصية الاستخراج |
| | | ق. تقديم عدد من الاختيارات الملائمة لاستعراض المعجم المختص وفقًا لأنظمة ترتيب مخصصة. | أن يتيح المعجم عدة خيارات لاستعراض المداخل، مثل: الترتيب الأبجدي، أو حسب الشيوع، أو التصنيف الموضوعي. | نسبة تفعيل أنظمة الترتيب المتاحة | مراجعة أنظمة العرض المتاحة |
| | | ر. استعمال المرشحات والفلاتر البحثية المتقدمة لتصفية نتائج البحث وإعادة تجميعها بطرق مختلفة. | أن يقدّم محرك البحث فلاتر متقدمة تتيح للمستخدمين تصفية نتائج البحث حسب الفئات المختلفة (مثل: المجال، والصرف، والدلالة). | نسبة تفعيل المرشحات المتاحة في تحسين نتائج البحث | مراجعة وظائف التصفية وتحليل الاستخدام |
| | | ش. عرض نتائج البحث بطرق مناسبة ومختصرة. | عرض نتائج البحث بطريقة مختصرة ومنسقة، ليحصل المستخدم على لمحة سريعة عن المعلومات دون الحاجة إلى التصفح المطول. | معدل رضا المستخدم عن عرض النتائج | استبيانات + مقابلات بعد البحث |
| | | ت. إمكانية التحكم والبحث في نتائج الاستعلام الخاصة بالمعجم المختص. | تمكين المستخدم من التحكم في نتائج بحثه من خلال تعديل الاستعلامات أو تصفية النتائج للوصول إلى المعلومات المطلوبة بسرعة. | نسبة مرونة التحكم في النتائج | اختبار عملي لخيارات التحكم في النتائج |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| **سابعًا: معايير الأدوات والتقنيات الإثرائية للمعجم المختص** | ١. توافر أدوات وتقنيات إثرائية للمعجم المختص | أ. توفر مساعد ذكي لتعليم الكتابة والقراءة والمفردات (المهارات اللغوية الاستقبالية والإنتاجية). | أن يحتوي المعجم على مساعد ذكي يساعد المستخدمين في تطوير مهارات القراءة والكتابة، بالإضافة إلى تعزيز المفردات، عن طريق تقديم تمارين تفاعلية أو نصائح. | توفر المساعد الذكي للكتابة. | مراجعة تقنية- اختبار عملي |
| | | ب. توفر أداة لتحليل النصوص المكتوبة وتصنيفها أو تميزها آليًا وفقًا لبيانات المعجم المختص. | أن يتيح المعجم أداة لتحليل النصوص المكتوبة وتصنيفها تلقائيًا بناءً على البيانات المعجمية؛ مما يساعد المستخدمين على فهم النصوص وتصنيفها وفقًا للمفاهيم اللغوية المدرجة في المعجم. | توفر أداة التحليل والتصنيف الآلي للنصوص. | مراجعة تقنية- اختبار عملي. |
| | | ج. توافر أدوات تساعد ذوي الاحتياجات الخاصة على استعمال المعجم المختص. | أن يتضمن المعجم أدوات تتيح لذوي الاحتياجات الخاصة استخدامها بسهولة، مثل: أدوات تكبير النصوص، أو قارئات الشاشة، أو التحكم الصوتي. | توفر وتفعيل أدوات المساعدة لذوي القدرات الخاصة | مراجعة تقنية - اختبار عملي |
| | | د. توافر خاصية الإطارات المنبثقة للتطبيقات الخاصة بإثراء المعجم المختص بمصادر خارجية. | أن يتيح المعجم خاصية الإطارات المنبثقة التي تربط المستخدم بمصادر خارجية؛ مما يمنحه إثراءً معرفيًا إضافيًا دون مغادرة الصفحة. | نسبة المداخل المربوطة بإطارات منبثقة إثرائية | اختبار عملي لوظائف الإطارات المنبثقة وربطها بمصادر خارجية |
| | | هـ. توظيف تقنيات مبتكرة وحديثة لإدارة البيانات المعجمية وتصفحها والبحث فيها. | أن يستخدم المعجم تقنيات حديثة مثل الذكاء الاصطناعي والتعلم الآلي لتحسين إدارة البيانات، وتسريع البحث، وتوفير تجربة تصفح سلسة وفعالة. | عدد التقنيات الحديثة المدمجة | مراجعة المنصة وتوثيق الوظائف المضافة |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعايير التفصيلية | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | و. استعمال أنظمة التوصية القائمة على الذكاء الاصطناعي لتقديم مقترحات المداخل المعجمية للمستخدم. | أن يستخدم المعجم تقنيات الذكاء الاصطناعي لتقديم توصيات مخصصة للمستخدم، بناءً على استعلاماته السابقة وسجل بحثه؛ مما يعزز من تجربة المستخدم. | نسبة دقة المقترحات المقدمة من توصيات الذكاء الاصطناعي | اختبار آلي للتوصيات + تحكيم بشري لعينة |
| | | ز. التوسيم اللغوي للنصوص والأمثلة المستعملة في المعجم المختص (الصرفي، والنحوي، والدلالي، والتداولي). | تصنيف وتوسيم النصوص والأمثلة في المعجم حسب عدة معايير عدة، مثل: الصرف، والنحو، والدلالة، والتداول؛ مما يساعد المستخدمين على فهم أعمق للعلاقات اللغوية. | نسبة النصوص والأمثلة الموسومة بشكل صحيح | مراجعة لغوية – تدقيق وسوم النصوص بعينة مرجعية وتحكيم خبير لغوي |
| | | ح. استعمال الرسوم البيانية والمشجرات لتنظيم بيانات المعجم المختص وتيسير التنقل بين أجزائه. | استخدام الرسوم البيانية والمشجرات لتوضيح العلاقات الدلالية والصرفية بين المداخل المعجمية؛ مما يسهل على المستخدمين التنقل بين المفاهيم المختلفة. | نسبة المداخل التي تحتوي على رسوم بيانية أو مشجرات. | مراجعة مخرجات النظام وتجربة التنقل بين المكونات |
| | | ط. توظيف تقنيات مبتكرة وحديثة لإدارة البيانات المعجمية وتصفحها والبحث فيها. | استخدام تقنيات حديثة، مثل: قواعد البيانات العلائقية أو الرسومية، والذكاء الاصطناعي، لتحسين إدارة البيانات؛ مما يتيح للمستخدمين تجربة بحث وتصفح سلسة وسريعة. | توفر تقنيات حديثة لإدارة البيانات المعجمية وتصفحها | مراجعة تقنية – اختبار عملي |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

## المراجع :


- القاسـمي، علـي محمـد. (٢٠١٩). علـم المصطلـح: أسسـه النظريـة وتطبيقاتـه العمليـة، بـيروت، مكتبـة لبنـان ناشـرون.

- Alberts, M. (2001). Lexicography versus terminography. *Lexikos, 11*, 1–12. https://doi.org/10.5788776-0-11/

- Fuertes-Olivera, P. A., & Tarp, S. (2014). *Theory and practice of specialised online dictionaries: Lexicography versus terminography* (Vol. 146). Walter de Gruyter.

- International Organization for Standardization. (2001). *ISO 15188: Project management guidelines for terminology standardization*. ISO.

- International Organization for Standardization. (2007). *ISO 860: Terminology work—Harmonization of concepts and terms.* ISO.

- International Organization for Standardization. (2008). *ISO 22128: Terminology products and services—Overview and guidance*. ISO.

- International Organization for Standardization. (2014). *ISO 241561-: Graphic notations for concept modelling in terminology work and its relationship with UML—Part 1: Guidelines for using UML notation in terminology work.* ISO.

- International Organization for Standardization. (2015a). *ISO 17100: Translation services—Requirements for translation services*. ISO.

- International Organization for Standardization. (2015b). *ISO 9000: Quality management systems—Fundamentals and vocabulary*. ISO.

- International Organization for Standardization. (2019a). *ISO 1087: Terminology work and terminology science—Vocabulary*. ISO.

- International Organization for Standardization. (2019b). *ISO 261621-: Management of terminology resources—Terminology databases—Part 1: Design*. ISO.

- International Organization for Standardization. (2019c). *ISO 30042: Management of terminology resources—TermBase eXchange (TBX)*. ISO.
- International Organization for Standardization. (2021). *ISO 126161-: Terminology work in support of multilingual communication—Part 1: Fundamentals of translation-oriented terminography*. ISO.
- International Organization for Standardization. (2022a). *ISO 126201-: Management of terminology resources—Data categories—Part 1: Specifications*. ISO.
- International Organization for Standardization. (2022b). *ISO 704: Terminology work—Principles and methods*. ISO.
- International Organization for Standardization. (2022c). *ISO SMART standardization proposal* [Unpublished standard]. ISO/TC 37.
- International Organization for Standardization. (2023). *ISO/FDIS 24183: Technical communication—Vocabulary* (Final Draft International Standard). ISO.
- Jackson, H. (2002). *Lexicography: An introduction*. Routledge.
- Ježek, E. (2016). *The lexicon: An introduction*. Oxford University Press.
- L›Homme, M.-C., & Cormier, M. (2014). Dictionaries and the digital revolution: A focus on users and lexical databases. *International Journal of Lexicography, 27*(4), 331–340. https://doi.org/10.1093/ijl/ecu019
- Morris, L. H. (2022). Howard Jackson (Ed.). The Bloomsbury handbook of lexicography. *Lexikos, 32*, 364–367.
- Romary, L. (2001, February). An abstract model for the representation of multilingual terminological data: TMF—Terminological Markup Framework. *In TAMA 2001—5th TermNet Symposium (pp. 1–9).* Antwerp, Belgium. https://inria.hal.science/inria-00100405
- Sterkenburg, P. G. J. van (Ed.). (2003). *A practical guide to lexicography*. John Benjamins.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً


- Trap-Jensen, L. (2018). Lexicography between NLP and linguistics: Aspects of theory and practice. In J. Čibej, V. Gorjanc, I. Kosem, & S. Krek (Eds.), *Proceedings of the 18th EURALEX International Congress*: *Lexicography in global contexts* (pp. 17–21). Ljubljana University Press.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data, 3, Article 160018*. https://doi.org/10.1038/sdata.2016.18

# الفصل الخامس

## معايير بناء المدونات اللُّغوية

- ملخص
- مقدمة
- مدخل تعريفي إلى بناء المدونات اللُّغوية
- مسوغات وضع المعايير في بناء المدونات اللُّغوية
- الدراسات السابقة
- نطاق المعايير وحدودها
- قائمة المعايير
- المراجع



هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

## ملخص

تتطرق هذه الوثيقة إلى معايير بناء المدونات، وتجدر الإشارة هنا إلى اختلاف عملية بناء المدونة عن توسيمها. حيث يُقصد بالتوسيم: إضافة معلومات ووسوم متنوعة إلى نصوص المدونة اللُّغوية؛ لأغراض مهمات المعالجة الحاسوبية. وعلى الرغم من تطابق بعض معايير المدونات الخام والمُوَّسمة -مثل معايير تخزين وحفظ المدونات- إلا أن الفرق الجوهري بينهما يكمن في تركيز هذه الوثيقة على المعايير التي ينبغي مراعاتها في جميع مراحل عملية بناء المدونات بشكل عام، دون التطرق إلى ما يتعلق بإضافة الوسوم المتنوعة إلى المدونة الخام، التي غالبًا ما تكون مصطنعةً وليست موادَّ لغويةً تُستخدَم في شكلها الطبيعي. وعادةً ما يتطلب التوسيم عملًا جماعيًّا وفهمًا أوسع للغة وتعقيداتها. وعلى ذلك، فإنه من المناسب التفريق في وثائق المعايير بين توسيم المدونات -مثل معايير تدريب المُوسِّمين، وإعداد أدلة التوسيم- وبناء المدونات، مثل: معايير اختيار عينات النصوص العشوائية، وتمثيل المدونات، وغيرها. ولذلك تقتصر هذه الوثيقة على سرد معايير بناء المدونات اللُّغوية التي تتضمن أهم المعايير الخاصة بآلية جَمْع المدونات اللُّغوية وتنظيمها وتخزينها. ويُساعد الالتزام بهذه المعايير على ضمان بناء مدونات لغوية عربية ذات جودة عالية وفقًا للمواصفات المعتبرة، وتبعًا للمنهج العلمي الخاص ببناء المدونات اللُّغوية وتطويرها لأغراض استعمالها في المهمات المختلفة لمعالجة اللُّغة الطبيعية. وقد رُوعي في هذه المعايير أن تكون مرنةً وشاملةً لجميع أنواع المدونات اللُّغوية، وغير مُقيَّدة للعمل بتقنيات معينة؛ وذلك استجابةً للتطوُّر المستمر في أساليب وتقنيات جَمْع المدونات وبنائها.

## مقدمة

شـهد العالـم في السـنوات الأخـيرة طفـرةً هائلـةً في مجـال الـذكاء الاصطناعـي، إذ باتـت الأنظمـة والتطبيقـات الحاسـوبية تقـترب شـيئًا فشـيئًا مـن محـاكاة الـذكاء البشـري، مكتسـبةً قـدراتٍ متقدمـةً في التعلُّم الـذاتي، والاسـتنتاج، واتخاذ القرار حتى في سـياقات غـير مبرمجـة سـلفًا، وبـدون تدخُّـل مباشـر مـن الإنسـان. وقـد تَصـدَّر هـذا المشـهد نـوعٌ جديـد مـن الـذكاء الاصطناعـي عُـرف بـالذكاء الاصطناعـي التوليـدي (GenAI)، الـذي يعتمـد علـى كميـات هائلـة مـن البيانـات للتدريـب، ليُنتـج أنظمـةً فائقـةَ الأداء، مـن أبرزهـا النمـاذج اللغويـة الضخمـة (Large Language Models – LLMs) مثـل ChatGPT وLLaMA وغيرهـا، والـتي أظهـرت قـدراتٍ مذهلـةً في فهـم اللغـة البشـرية وإنتاجها بصـورة أقـرب مـا تكـون إلى التعبـير الإنسـاني الطبيعـي. هـذا التطـور اللافـت سـلّط الضـوء علـى الحاجـة المُلحّـة إلى بيانـات لغويـة رقميـة عاليـة الجـودة، إذ أصبحـت هـذه البيانـات بمثابـة الوقـود الحيـوي الـذي تُبـنى عليـه كفـاءة هـذه النمـاذج، ويُقـاس عـن طريقهـا مسـتوى أدائهـا، ودقّـة تطبيقاتهـا في مجـالات عـدة مـن معالجـة اللغـة الطبيعيـة.

**ومـن أهـم المـوارد اللُّغويـة الرقميـة في هـذا السـياق** مـا يُعـرف بـ(المدونـات اللُّغويـة) (Corpora)، الـتي تُعَـدُّ البنيـةَ الرئيسـة الـتي تُبـنى عليهـا النمـاذج اللُّغويـة الضخمـة، كمـا تؤثـر طريقـة بنـاء هـذه المدونـات ومـدى دقتهـا في جـودة المخرجـات اللُّغويـة لنمـاذج الـذكاء الاصطناعـي، فضـلًا عـن أهميتهـا في الدراسـات اللُّغويـة واللسـانية الـتي تتخـذ مـن المدونـات أداة للتحليـل والرصـد، واسـتنباط الظواهـر اللغويـة

**تهدف هـذه الوثيقـة** إلى بنـاء قائمـة تتضمـن عـددًا مـن المعايـير القياسـية الـتي يمكـن الاعتمـاد عليهـا عنـد جَمْـع المدونـة الخـام وبنائهـا، وتسـتهل الوثيقـة بمدخـل إلى بنـاء المدونـات الخـام واسـتعراض أهـم المفاهيـم المسـتخدمة فيهـا، ثـم يتبـع ذلـك مدخـل تعريفـي بمعايـير بنـاء المدونـات اللُّغويـة وحدودهـا،

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

مع ذكر أهم دوافع وَضْع هذه المعايير واقتراح بعض الطرق لتطبيقها. وفي إطار التأصيل العلمي، تستعرض الوثيقة أبرز الأدبيات السابقة التي تناولت معايير بناء المدونات اللغوية، وذلك لوضع هذه الوثيقة في سياقها المعرفي والتطبيقي. وتُختتم الوثيقة بعرض قائمة تفصيلية للمعايير المقترحة لبناء المدونات الخام، بما يُسهم في تقديم إطار عمل متكامل، يمكن الاستناد إليه لضمان جودة المدونات اللغوية وتوظيفها بكفاءة في مجالات متعددة.

## مدخل تعريفي:

يتناول هذا القسم المفاهيم الرئيسة المرتبطة بالمدونات اللغوية؛ بما في ذلك تعريفها، والفرق بين المدونات الخام والمدونات الموسمة، وأنواع المدونات وفقًا لمعايير متعددة، ومناقشة اعتبار شبكة الإنترنت مصدرًا لجمع المواد اللُّغوية، مع ما يثيره هذا الخيار من فرص وتحديات تتعلق بالتنوُّع، والجودة، والموثوقية.

## تعريف المدونة اللُّغوية:

يُقصَد بالمدونة اللُّغوية (Corpus) في سياق هذه الوثيقة: مجموعة من المواد اللُّغوية الأصيلة في وسائط متعددة، التي تُحوَّل غالبًا إلى نصوص رقمية؛ لأغراض النمذجة والمعالجة الحاسوبية الآلية للغة باستخدام الحاسوب (Zeroual & Lakhouaja, 2018; Wynne, 2005).

ومن المصطلحات الأخرى التي تُسمَّى بها المدونة اللُّغوية: "الذخيرة اللُّغوية" أو "المكنز" (العصيمي، ٢٠١٥: ٢٠)، واخترنا مصطلح المدونة في هذه الوثيقة؛ لأنه الأكثر استعمالًا وشيوعًا. وقد تجمع المدونات اللُّغوية في مجموعة من الأوراق أو الكتب بطرق تقليدية؛ لأغراض البحوث اللُّغوية المتنوعة، ولا يُعد هذا النوع من المدونات معتبرًا في سياق اللسانيات الحاسوبية إلا بعد تحويلها إلى نصوص رقمية يمكن التعامل معها ونمذجتها في مهمات المعالجة الحاسوبية المتعددة.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً



ويذكـر (Wynne 2005, 22) في تعريـف المدونـة اللُّغويـة: أنهـا مجموعـة مـن "المقاطـع" للنصـوص اللُّغويـة محفوظـة في صيـغ إلكترونيـة، اختِـيرت وفقًـا لمعايـير خارجيـة؛ لتُمثِّـل بقـدر الإمـكان لغـةً أو لهجـةً مُحـدَّدة، ويمكـن اعتمادهـا مصـدرًا لبيانـات البحـوث اللُّغويـة. ويسـتخدم Wynne كلمـة "مقاطـع"؛ لأن بعـض المدونـات تقتصـر علـى جَمْـع عينـات وأجـزاء مـن النصـوص بـدلًا مـن جمـع نصـوص كاملـة أو تفريـغ نـصي لأصـوات تمثـل سـياقات متكاملـة. ويُوضِّـح Wynne (٢٠٠٥) في سـياق تعريـف المدونـة، أنـه لا يمكـن اعتبـار أي نصـوص إلكترونيـة مُجمَّعـة مدونـةً لغويـةً. فعلـى سـبيل المثـال، لا تعتـبر النصـوص الخاصـة بالاستشـهادات والمراجـع والاقتباسـات أو النصـوص الأرشـيفية مدونـات لغويـة وفقًـا للتعريـف الخـاص بهـا. ويضيـف كذلـك أن نصـوص شـبكة الإنترنـت لا ينطبـق عليهـا تعريـف المدونـة اللُّغويـة، وهـو رأي يُخالفـه فيـه عـدد مـن الباحثـين، مـن أبرزهـم Nakov (٢٠١٤)، ويعـزو ذلـك Wynne إلى عـدد مـن الإشـكاليات، مـن أبرزهـا: غمـوض حـدود النصـوص الإلكترونيـة، وعـدم وضـوح حجمهـا الفعلـي، والتغيـر الدائـم لمحتواهـا، فضـلًا عـن أن هـذه النصـوص لـم تُنشـأ في الأصـل لأغـراض لغويـة بحثيـة؛ لكنـه أكَّـد في الوقـت نفسـه أن شـبكة الإنترنـت تُعَـدُّ مصـدرًا غنيًّـا ومتنوعًـا لباحـثي اللغـة، ويمكـن توظيفهـا بفاعليـة في جمـع وبنـاء المدونـات اللغويـة لأهـداف بحثيـة وتطبيقيـة متنوعـة.

ومـن جهـة أخـرى، يعتـبر Meyer (٢٠٠٣) نصـوص الإنترنـت مدونـة لغويـة، علـى الرغـم مـن غمـوض حجمهـا وتفـاوت نوعيـة نصوصهـا. ويؤيـد هـذا الـرأي Nakov (٢٠١٤) الـذي يؤكـد أن الإنترنـت هـي أكبـر مدونـة لغويـة. وعلـى الرغـم مـن وجـود مثـل هـذه الاختلافـات في اعتبـار شـبكة الويـب مدونـة لغويـة إلا أن الجميـع يؤكـد أن النصـوص المتاحـة علـى شـبكة الإنترنـت تُعَـدُّ مصـدرًا مهمًّـا لبنـاء المدونـات، وجَمْـع البيانـات اللُّغويـة الـتي يمكـن توظيفهـا بطـرق متعـددة في الدراسـات اللُّغويـة ومعالجـة اللُّغـة الطبيعيـة.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

## أنواع المدونات وتقسيمها:

يمكـن تصنيـف المدونـات وفقًـا لاعتبـارات متعـددة، مثـل: تقسـيمها **وفقًـا للغـرض الـذي تُسـتعمَل فيـه**، حيـث يمكـن اسـتعمالها مدونـةً عامـةً أو متخصِّصـةً، كمـا يمكـن تقسـيمها **وفقًـا لطبيعـة المحتـوى ونـوع اللُّغـة** إلى مدونـات مكتوبـة أو منطوقـة، ويمكـن تقسـيمها كذلـك **وفقًـا للبيانـات الوصفيـة المصاحبـة لها** إلى فئتـين رئيسـتين **همـا:**

- **المدونات الخام**، وهي محور هذه الوثيقة.
- **المدونات الموسمة أو الموصوفة**، وخُصِّص لها وثيقة منفصلة.

وتقتصـر الفئـة الأولى علـى جَمْـع المـواد اللُّغويـة بشـكلها الخـام الطبيعـي، وبعـد ذلـك تخزينهـا في قوالـب وصيـغ رقميـة في الحاسـوب بطـرق متعـددة، أمـا الفئـة الثانيـة فتُضـاف فيهـا إلى مـواد المدونـة الخـام أنـواعٌ مختلفـةٌ مـن البيانـات الوصفيـة والوسـوم اللُّغويـة بطـرق وأسـاليب متعـددة، عـادةً مـا تتطلَّـب معرفـةً لغويـةً أعمـق، وتدريبًـا خاصًّـا، وتسـهم هـذه الوسـوم في تعـدُّد اسـتخدامات المدونـة اللُّغويـة وزيـادة احتمـالات الإفـادة منهـا عنـد اسـتعمالها في تطبيقـات المعالجـة الآليـة للغـة؛ إذ يتمكَّـن النمـوذج اللُّغـوي والخوارزميـات الحاسـوبية عنـد معالجـة المدونـات الموسـمة مـن الوصـول إلى رؤى لغويـة أعمـق وتحليـل أدق، ويؤثـر هـذا مباشـرة في تحسـين جـودة مخرجـات تطبيقـات الـذكاء الاصطناعـي واللسـانيات الحاسـوبية. وعلـى ذلـك، فـإن الفـرق بـين المدونـة الخـام والمدونـة الموسـمة، يمكـن اختصـاره في أن الأولى هـي نتيجـة اسـتعمال لغـوي طبيعـي، في حـين الأخـرى تتطلَّـب إضافـةَ معلومـات لغويـة إضافيـة لا تُسـتخدَم بشـكل طبيعـي؛ بـل تهـدف إلى توضيـح المدونـة الخـام وإثرائهـا بمـواد لغويـة لا تظهـر في صورتهـا الطبيعيـة.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً



ونظـرًا لتعـدُّد اسـتخدامات المدونـات اللُّغويـة واختـلاف طـرق بنائهـا؛ فـإن لهـا أنواعًـا متعـددةً، تختلـف وفقًـا لاعتبارات متعـددة، مثل: العمـوم، والخصوص، والزمـن. وقد يصعـب تصنيـف المدونـة بشـكل كامـل في فئـة محـددة، ولكـن لـكل مدونـة مواصفات وخصائـص معينـة تُميِّزهـا عـن غيرهـا، **ومـن أهـم طـرق تصنيـف المدونـات مـا يـأتي:**

- **تعدُّد اللُّغات:**

— **المدونـات أحاديـة اللُّغـة (Monolingual):** وهـي النـوع الأكـثر شـيوعًا مـن المدونـات، وتكـون بلغـة واحـدة، مثـل: العربيـة، أو الإنجليزيـة، أو اليابانيـة.

— **المدونـات المتوازيـة (Parallel):** وهـي مدونـة مـن أكـثر مـن لغـة، تتألـف مـن نصـوص متوازيـة مكتوبـة بلغـات مختلفـة، حيـث تكـون المـواد اللُّغويـة في كل لغـة عبـارة عـن ترجمـة عـن الأخـرى.

- **النوع اللُّغوي:**

— **المدونـة العامـة:** وهـي مدونـة بُنيـت لأغـراض عامـة، ولا تغطـي نصوصًـا أو موضوعـاتٍ معينـةً، مثـل المدونـة العربيـة.

— **المدونة الخاصة:** وهي مدونة بُنيت من موضوعات محددة، مثل النصوص الدينيـة أو السياسـية، أو مـن مناطـق جغرافيـة معينـة، أو لمُؤلِّـف معـين، أو مختصـة بنـصٍّ مُحـدّد، مثـل مدونـة القـرآن الكريـم.

— **المدونـات المقارنـة Comparable:** وهـي مدونـة مـن نصـوص متشـابهة مـن حيـث المحتـوى؛ لغـرض مقارنـة الاسـتعمالات اللُّغويـة المختلفـة بينهـا، مثـل: المدونـات اللُّغويـة الـتي تجمـع الرسـائل الجامعيـة في موضوعـات محـددة (مثـل: علـوم الحاسـوب)، ومقارنـة أسـاليب الكتابـة بـين المتحدثـين باللُّغـة كلغـةٍ أم ولغـةٍ ثانيـة.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

- **نوع محتوى الوسائط اللُّغوية:**

— **المدونات النصية:** وهو النوع الغالب من المدونات اللُّغوية، وتكون في أساسها موادَّ لغويةً مكتوبةً، مثل: المقالات الصحفية، والكتُب، والرسائل الجامعية.

— **المدونات الصوتية:** وهي المدونات التي تُجمع عن طريق تسجيل أصوات المتحدثين، وقد يتبعها تفريغ الكلام المنطوق إلى مكتوب. تشمل المدونات النصية: المقابلات التلفزيونية، والتسجيلات الهاتفية، والتعليقات الإذاعية، والمحادثات اليومية.

- **مواصفات الزمن:**

— **المدونات التزامنية Synchronic:** تغطي فترةً زمنيةً مُحدَّدةً، ولا تشمل فترات مختلفة جوهريًّا في استعمالات اللُّغة، وغالبًا ما تُجمع نصوصها في سنوات متقاربة، حيث لا يظهر تباين ملحوظ في خصائص اللغة أو طرائق استخدامها.

— **المدونات التاريخية Diachronic corpus:** تغطي فترات زمنية مختلفة؛ لغرض مقارنة الاستعمالات اللُّغوية، مثل: المدونة العربية التي تغطي فترة ما قبل الإسلام إلى العصر الحديث. وقد تكون الفترات الزمنية متتابعةً أو متفرقةً حسب الهدف من جَمْع المدونة.

- **مواصفات التغيُّر والثبات عبر الزمن:**

— **المدونة الثابتة Static corpus:** وتُسمَّى كذلك (المدونة المرجعية reference corpus)، وهي مدونة مكتملة البناء لا يتغير محتواها؛ إذ إن الهدف من بنائها لا يتغير، ولا تتغير نتائج التحليل اللُّغوي عليها. ومعظم المدونات اللُّغوية هي مدونات ثابتة.

— **مدونة المتابعة monitor:** بعكس المدونة الثابتة، يتغير محتوى مدونة المتابعة عبر الوقت، وتُستخدَم مدونة المتابعة في متابعة ورصد التغيُّرات اللُّغوية، حيث يتم تحديث محتواها بشكل دوري.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً



- **الوسوم اللُّغوية:**

— **المدونة الخام:** وهي مدونة مكونة من مواد لغوية بشكلها الطبيعي، مثل: المقالات الصحفية، والقصائد الشعرية، وذلك دون إضافات جوهرية عليها. باستثناء بعض البيانات الوصفية، مثل: وقت الجمع، ومصدره، وغيره. وهي محور هذا الفصل.

— **المدونة المُوسَّمة:** هي مدونة يُضاف إليها معلومات لغوية جوهرية لا توجد في المدونة الخام، مثل: إضافة أقسام الكلام إلى المدونة. وهي محور اهتمام فصل "معايير توسيم المدونات اللُّغوية".

**مصادر بناء المدونات اللُّغوية العربية:**

مع تطوُّر شبكة الإنترنت وظهور الشبكة العالمية (World Wide Web) عام ١٩٩١ على يد (تيم بيرنرز لي)، تحوَّلت الإنترنت إلى منصة مفتوحة على مستوى العالم، تتيح الوصول إلى المحتوى والمساهمة في إنتاجه بسهولة غير مسبوقة. وقد أسهم هذا التحول، -ولا سيما مع بروز مفهوم "الويب ٢,٠" الذي يُعزِّز مشاركة المستخدمين في صناعة المحتوى- في تضخم هائل لحجم البيانات اللُّغوية المتوفرة على الشبكة، وتنوُّعها بشكل ملحوظ. نتيجة لذلك، أصبح الإنترنت المصدر الرئيس لبناء المدونات اللغوية، لما يتزخر به من نصوص ضخمة ومتنوعة، يُمكن الوصول إليها وتحديثها بشكل لحظي. وتشير الدراسات إلى أن أكثر من ٥١٪ من المدونات بُنيت جزئيًّا أو كليًّا من محتوى الويب (Zeroual & Lakhouaja ,2018).

**مزايا استعمال شبكة الإنترنت في بناء المدونات اللُّغوية العربية:**

**١- ضخامة المحتوى اللُّغوي وتنوُّعه:**

- **بيانات غير محدودة:** يحتوي الويب على كمية ضخمة من النصوص من أنواع مختلفة؛ بما في ذلك المقالات الإخبارية، والمدونات، والمنتديات، ومنشورات وسائل التواصل الاجتماعي، والأبحاث الأكاديمية، وغيرها. وهذا التنوُّع ضروريٌّ لبناء مدونات متوازنة وتمثيلية.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

- **مـوارد متعـددة اللُّغـات:** تُوفِّـر المواقـع الإلكترونيـة بيانـات بلغـاتٍ عـدة؛ ممـا يتيـح فرصًـا قيّمـة للدراسـات متعـددة اللُّغـات واللُّغويـات المقارنـة.

**٢- سهولة الوصول وتحديث البيانات:**

- **سـهولة الوصـول:** يمكـن للباحثـين الوصـول بسـهولة إلى كميـة هائلـة مـن البيانـات النصيـة دون الحاجـة إلى جَمْعهـا بشـكل فعلـي؛ وهـو مـا يجعـل الويـب مصـدرًا فاعِـلًا لبنـاء المدونـات.
- **محتـوى حديـث وديناميكـي:** يتـم تحديـث الويـب باسـتمرار بمحتـوى جديـد؛ وهـو مـا يسـمح للمدونـات بتضمـين أحـدث اتجاهـات اسـتخدام اللُّغـة والبيانـات في الوقـت الفعلـي.

**٣- تطوُّر الأدوات والتقنيات المُتخصِّصة:**

- **أدوات اسـتخراج البيانـات مـن الويـب:** تطـوَّرت الأدوات والبرمجيات لاسـتخراج المحتـوى النـصي بكفـاءة مـن المواقـع الإلكترونيـة.
- **تطـوُّر واجهـات برمجـة التطبيقـات (APIS) وقواعـد البيانـات:** تُوفِّـر العديـد مـن المنصـات واجهـات برمجـة تطبيقـات تتيـح للباحثـين الوصـول المباشـر إلى كـمٍّ هائـل مـن البيانـات اللُّغويـة بطريقـة منهجيـة وفعَّالـة.
- **زيـادة القـدرة الحاسـوبية:** سـاعدت الطفـرات المتسـارعة في قـدرات المعالجـة والتخزيـن الحاسـوبي علـى التعامـل مـع الكميـات الضخمـة مـن البيانـات المتاحـة عـبر الويـب، ممـا جعـل مـن الممكـن تحليـل محتـوى الإنترنـت المتنـامي بكفـاءة عاليـة.
- **تطـوُّر أدوات الزحف والاسـتخراج:** تتيـح أدوات الزحـف (Crawling) والاسـتخراج التلقائيـة جَمْـع بيانـات الويـب بكفـاءة؛ ممـا يجعـل العمليـة قابلـة للتطويـر علـى نطـاق واسـع.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً



- **أدوات معالجة اللُّغة الطبيعية NLP:** حسَّنت الأدوات والخوارزميات المتقدمة لمعالجة اللُّغة الطبيعية من معالجة وتنظيف بيانات الويب؛ مما زاد من قابليتها للاستخدام في البحث اللُّغوي.

**زواحف الويب (Web Crawlers):**

هي برامج حاسوبية مُتخصِّصة تُستخدَم لتصفُّح صفحات الويب واستخلاص المعلومات منها بطريقة آلية ومنهجية. وتُعَدُّ زواحف الويب من الأدوات الأساسية في جمع البيانات من الإنترنت، حيث تعتمد عليها محركات البحث في فهرسة الصفحات، كما تُستخدم على نطاق واسع في تجميع النصوص لبناء المدونات اللغوية. تعمل على زيارة الصفحات واستخلاص النصوص منها، ثم الانتقال إلى الصفحات المرتبطة بها عبر الروابط بشكل تكراري؛ مما يسمح بجَمْع كميات كبيرة من النصوص في وقت قصير.

**بناء المدونات الرقمية من المصادر غير الرقمية:**

إضافةً إلى الويب، تُعَدُّ المصادر اللُّغوية غير الرقمية -مثل الكتب المُصوَّرة- من أهم الموارد لبناء المدونات، ولا سيّما في اللغة العربية فيما يتعلّق بالمؤلَّفات التاريخية، حيث إن نسبة المحتوى العربي على الشبكة العنكبوتية تتراوح ما بين ٠.٥٪ إلى ٣٪ (١)، وذلك على الرغم من أنها من أكثر اللُّغات من حيث عدد المتحدثين. وقد انتشرت الكتب المصورة انتشارًا واسعًا، حيث يتجاوز عددها أكثر من ١٠٠ ألف كتابٍ متاح على الشبكة. **ويُمكن استخراج النصوص من هذه الكتب بطريقتين:**

١- **التعرُّف الضوئي على الحروف (OCR):** تُستخدَم هذه التقنية لتحويل الصور النصية إلى نصوص رقمية، إلا أن اللُّغة العربية تواجه تحديات في هذا المجال بسبب تعقيدات الخط العربي وتشكيل الحروف. وعلى الرغم من وجود أدوات مثل PyTesseract، إلا أنها لا تزال تحتاج إلى تحسينات للتعامل مع النصوص العربية بكفاءة عالية؛ وهو ما يؤثر على جودة النص المُستخرَج.

(1) https://www.unescwa.org/
https://www.statista.com/statistics/262946/most-common-languages-on-the-internet/

٢- **اسـتخراج النصـوص مـن الملفـات الرقميـة:** في حالـة توفُّـر النصـوص بصيغـة رقميـة مثـل ملفـات Word، يمكـن اسـتخراج النصـوص برمجيًّـا بسـهولة عـن طريـق اسـتخدام لغـات البرمجـة مثـل Python ومكتباتهـا. أمـا في حالـة ملفـات PDF، فـإن اسـتخراج النصـوص العربيـة يواجـه تحديـات تقنيـة بسـبب مشـكلات التشـفير النـصي في هـذه الملفـات؛ ممـا يجعـل العمليـة أكـثر تعقيـدًا، ويتطلـب أدوات متخصِّصـة للتعامـل مـع التشـفير العـربي.

**تحديات جَمْع البيانات من المصادر غير الرقمية:**

- ضمـان جـودة النصـوص المُسـتخرَجة: قـد تحتـوي النصـوص المُسـتخرَجة علـى أخطـاء نتيجـة لعـدم دقـة تقنيـات التعـرُّف الضـوئي علـى الحـروف أو مشـكلات التشـفير؛ ممـا يسـتدعي عمليـات تنقيـح وتصحيـح إضافيـة.
- الضوضـاء النصيـة: وجـود رمـوز أو تنسـيقات غـير مرغـوب فيهـا في النصـوص المُسـتخرَجة؛ ممـا يتطلـب خطـوات تنظيـف ومعالجـة مسـبقة قبـل اسـتخدامها في بنـاء المدونـة.
- حقـوق الملكيـة الفكريـة: ينبغـي مراعـاة الجوانـب القانونيـة عنـد اسـتخدام الكتـب المصـورة والمـواد المحميـة بحقـوق النشـر في بنـاء المدونـات اللُّغويـة.

## مُسوِّغات بناء المعايير:

شـهدت السـنوات الأخـيرة تصاعـدًا سـريعًا في وتـيرة الأبحـاث والتطبيقـات المعتمـدة على الـذكاء الاصطناعـي ومعالجـة اللُّغـة الطبيعيـة، ولا سـيما مـا يتصـل باللُّغـة العربيـة. وفي ظـلِّ هـذا التطـوُّر، ازدادت التحديـات المرتبطـة بجَمْـع البيانـات اللُّغويـة وتنقيحهـا وتنسـيقها، إذ كثـيرًا مـا نجـد مشـاريع متشـابهة تهـدر جهـدًا كبـيرًا في إعـادة العمـل مـن الصفـر أو في معالجـة نصـوص خـام تفتقـر إلى التوثيـق المنهـجي؛ وذلـك لغيـاب معايـير ومواصفـات متفـق عليهـا تُنظِّـم عمليـة بنـاء المدونـات اللُّغويـة. وفي الوقـت ذاتـه، تُطالـب مؤسسـات بحثيـة

مختلفة -سواء محلية أو دولية- بضرورة الارتقاء بجودة الموارد اللُّغوية الرقمية العربية وتحسينها؛ الأمر الذي لا يمكن بلوغه إلا بتوحيد الرؤى واعتماد معايير واضحة ومتفق عليها. من هنا تأتي أهمية وجود معايير دقيقة تُبيِّن الخطوات الرئيسة في اختيار مصادر النصوص، وتحديد ضوابط جَمْعها، وتوثيقها بشكل يُسهِّل مشاركتها مع الباحثين وتكرار استخدامها في مهمات لغوية متعددة. إنَّ توحيد المعايير يؤثر مباشرةً في مدى دقة النماذج الحاسوبية اللُّغوية وإمكانية تعميم نتائجها، ويُعزِّز فُرص التعاون بين مراكز البحث والشركات التقنية على مستوى العالم العربي.

## النتائج السلبية لغياب المعايير لبناء المدونات اللُّغوية:

في غياب معايير واضحة ومُحدَّدة، تتعدَّد السلبيات التي تمسُّ جودة المدونات اللُّغوية وقابليتها للاستخدام. فعلى سبيل المثال، قد تنشأ أخطاء فادحة في التصنيف أو في نسبة النصوص إلى فئاتها الصحيحة، وذلك عندما تعتمد الفِرَق البحثية على طُرق جمع متفاوتة أو اجتهادات شخصية في تنقية البيانات. كما يصعُب تقييم المدونات أو مقارنتها بمدونات أخرى إذا كانت معايير البناء مختلفةً جذريًّا من مشروع إلى آخر. أضف إلى ذلك أن غياب سياسة واضحة للتعامل مع حقوق الملكية الفكرية يُعرِّض مشاريع بناء المدونات إلى تحديات قانونية وأخلاقية، فضلًا عن احتمال حَجْب تلك المشاريع أو إيقافها. ويؤدي ذلك كله إلى هدر الجهد والوقت والموارد، حيث يجد الباحثون أنفسَهم مضطرين لإعادة هيكلة البيانات أو تصحيح الأخطاء الناتجة عن عملية جَمْع أو معالجة غير مُنظَّمة للنصوص. وبسبب هذه العشوائية، تتسع فجوة التعاون والتبادل بين المؤسسات والأفراد، وينحصر نطاق الاستفادة من المدونات في بيئة محدودة بدلًا من إتاحتها للباحثين في مجالات متنوعة، تشمل: اللسانيات الحاسوبية، وعلم اللُّغة الاجتماعي، وتحليل الرأي وغير ذلك من التطبيقات اللُّغوية المتطورة.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

## المبادئ المؤسّسة لهذه المعايير:

تنطلق هذه المعايير من جملة من المبادئ الرئيسة التي تُوجِّه عمليات بناء المدونات اللُّغوية، وتضمن قابليتها للاستخدام والتشارك في مشاريع متنوعة، **وأهم هذه المبادئ:**

١- **الشمولية والتمثيل:** تتطلب عملية بناء المدونات الحرصَ على تمثيل أكبر قدر ممكن من التنوع اللُّغوي في مضامين النصوص، سواء من حيث الموضوعات والمجالات المعرفية أو من حيث اللهجات والتراكيب اللُّغوية المختلفة. ويُفترَض أن تغطي المدونة عينةً وافيةً من الاستعمالات اللُّغوية المعاصرة أو التاريخية، بما يعكس واقعَ اللُّغة وثراءَها.

٢- **الاتساق والقابلية للتكرار:** يهدف وضع المعايير إلى ضمان اتباع الخطوات نفسها في جَمْع ومعالجة المواد اللُّغوية في أي مشروعٍ مشابه، إذ تتحدد بدقة طبيعة البيانات المطلوب جَمْعُها، وطريقة استخراجها، وآليات توثيقها؛ الأمر الذي يُسهِّل على أي فريق بحثي آخر إعادة تنفيذ العملية أو تحديثها تبعًا لحاجات جديدة.

٣- **الانفتاح والشفافية:** إذ يُفترض توثيق كل مرحلة من مراحل بناء المدونة بدقة، بما يشمل المصادر التي جُلبت منها النصوص، وأدوات الجَمْع، وطُرق التنظيف والتحقُّق من الجودة. وتُمكِّن هذه الشفافية الباحثين من فَهْم كيفية تكوين المدونة وإمكانية إعادة استخدامها، والتحقُّق من مصداقيتها.

٤- **القابلية للاستخدام والتشارك:** إن تعزيز الفائدة العملية للمدونة يتطلب توفيرها بصيغ متعدِّدة، مثل: ملفات بصيغة XML أو JSON وترميز UTF-٨، وتوفير وثائق توضح بنية البيانات والحقول الوصفية. وبذلك يسهل دمجها في برامج وأدوات معالجة اللُّغة الطبيعية، ويُعزز فرص التعاون البحثي بين المؤسسات الأكاديمية والصناعية.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً



٥- **الامتثـال القانـوني والأخـلاقي:** لا يخفـى مـا للملكيـة الفكريـة مـن أهميـة؛ إذ ينبغـي الحـرص علـى توثيـق حقـوق المُؤلِّفـين والناشـرين حـين تُـدرَج نصوصُهـم ضمـن المدونـة، والتأكـد مـن قانونيـة إعـادة نشـر البيانـات اللُّغويـة أو إتاحتهـا للباحثـين. وتشـمل المسـؤوليات الأخلاقيـة ضمـان عـدم انتهـاك النصـوص خصوصيـة الأفـراد، وخلـوّ هـذه النصـوص مـن المـواد الـتي تُحـرِّض علـى الكراهيـة، أو تخُالـف القواعـد المتفـق عليهـا في الأوسـاط العلميـة.

**أمثلة عملية على تطبيق هذه المبادئ:**

مـن أبـرز الأمثلـة علـى تجلّـي هـذه المبـادئ في بنـاء المدونـات العربيـة أو العالميـة، تضمـين خطـوات التنظيـف المعياريـة في صميـم منهجيـة الجَمْـع، حيـث تمـرُّ كلُّ الوثائـق أو المقاطـع الصوتيـة بمرحلـة فلـترة دقيقـة لإزالـة النصـوص المكـررة والعناصـر غـير اللُّغويـة كالإعلانـات والرمـوز البرمجيـة. يُضـاف إلى ذلـك ضـرورة تسـجيل أي معطيـات خاصـة بالمصـدر؛ كتاريـخ النشـر، والموقـع الجغـرافي، وفئـة النـص (أدبي، إخبـاري، أكاديـمي... إلـخ). ويسـاعد هـذا الوضـوح في تمكـين الباحثـين مـن إعـادة اختيـار عينـات محـددة لأغـراض دراسـية أخـرى أو التركـيز على تحليـل فـترة زمنيـة مُعيَّنـة. كذلـك، حـين يجـري تحديـث المدونـة دوريًّـا -كمـا في مدونـات المتابعـة- فـلا بـد مـن إصـدار نُسـخ مرفقـة بتوثيـق التعديـلات الجديـدة؛ ممـا يُـبرز مبـدأ الاتسـاق، ويُسـهِّل توظيـف الإصـدارات المختلفـة في مهمـات مقارنـة التطـوُّر اللُّغـوي أو دراسـة التغـيرات المصطلحيـة.

وفي المحصّلـة؛ فـإن الالـتزام بهـذه المبـادئ يُهـيِّئ سُـبُلًا أفضـل للاسـتفادة مـن المدونـات اللُّغويـة في الأبحـاث والتطبيقـات المختلفـة؛ إذ يُكسـبها جـودةً عاليـةً في بنائهـا واسـتدامةً في بقائهـا، ويجعـل منهـا مرجعًـا موثوقًـا في خدمـة الدراسـات اللسـانية الحاسـوبية، وتطويـر نمـاذج الـذكاء الاصطناعـي القائـم على اللُّغـة العربيـة، والتطبيقـات العلميـة والتجاريـة الـتي تسـتند إلى فَهْـم عميـق للنصـوص.

**تداعيات غياب المعايير على موثوقية المدونات اللغوية، وتكامل الجهود البحثية:**

**إن غياب المعايير يؤدي إلى:**

- تفاوت جودة المدونات اللُّغوية؛ مما يؤثر سلبًا على دقة النماذج والتطبيقات المبنية عليها.
- صعوبة التشارك والتكامل بين الباحثين والمطورين؛ نظرًا لاختلاف الأساليب والمنهجيات في جَمْع وبناء المدونات.
- زيادة الجهود والتكاليف؛ نتيجة إعادة العمل أو تصحيحه بسبب عدم اتباع معايير محددة.
- وتهدف هذه الوثيقة إلى صياغة عددٍ من المعايير القياسية التي يمكن الاعتماد عليها عند جَمْع وبناء المدونة الخام؛ وذلك من أجل:
- تحسين جودة البيانات اللُّغوية من خلال ضمان دقة المحتوى المُجمّع وموثوقيته.
- تسهيل عمليات المعالجة والتوسيم اللاحقة، بفضل تنظيم البيانات بشكل منهجي ومدروس.
- تعزيز التعاون والتكامل، عن طريق توفير أُسس مشتركة يمكن للباحثين الاعتماد عليها وتطويرها.

**أهمية وَضْع المعايير:**

تكمن أهمية وَضْع المعايير في تعزيز جودة المدونات اللغوية وموثوقيتها، ويتجلى ذلك في عدد من الجوانب، **أبرزها:**

- ضمان التمثيل اللُّغوي الشامل؛ حيث تضمن المعايير جَمْع نصوص تُمثِّل مختلف الأنماط، والأساليب، والحِقب الزمنية، واللَّهجات.
- تقليل التحيُّز والضوضاء، عن طريق تحديد مصادر موثوقة، وتطبيق إجراءات التنظيف والمعالجة الدقيقة.

- تسـهيل الاسـتخدام وإعـادة الاسـتخدام؛ فالمدونـات المبنيـة وفقًـا للمعايـير تكـون أكـثر قابليـةً للمشـاركة، وأسـهل توظيفًـا في سـياقات بحثيـة وتطبيقيـة متنوّعـة.
- مواكبـة التطـوُّرات التقنيـة، عـبر اعتمـاد معايـير تتوافـق مـع أحـدث التقنيـات والأدوات في مجـال معالجـة اللُّغـة الطبيعيـة.

## الدراسات السابقة:

يعـرض هـذا القسـم أهـم الأعمـال والدراسـات والتجـارب السـابقة حـول بنـاء المدونـات اللُّغويـة، وخصوصًـا الـتي تُركِّـز علـى بنـاء المدونـات العربيـة، مـع إشـارة إلى أهـم المدونـات العربيـة الـتي نُشـرت في السـنوات الأخـيرة، وتفاصيـل مواصفاتهـا ومعايـير بنائهـا.

بـدأت العنايـة بجَمْـع المدونـات اللُّغويـة منـذ ظهـور الحاسـوب وتطـوُّر اللسـانيات الحاسـوبية في سـتينيات القـرن المـاضي، وتطـوَّرت بعـد ذلـك طـرق ومناهـج بنـاء المدونـات اللُّغويـة، حـتى ظهـرت الحاجـة إلى وجـود معايـير مرجعيـة لضبـط عمليـة بنـاء المدونـات اللُّغويـة؛ لتحقيـق أكـبر فائـدة منهـا في المهمـات المختلفـة لمعالجـة اللُّغات. وذكـر Pustejovsky & Stubbs, (٢٠١٢) وIde (١٩٩٨) أن بدايـة جَمْـع المدونـات اللُّغويـة كانـت في الخمسـينيات مـن القـرن المـاضي، وكانـت تقتصـر علـى بنـاء مدونـات لغويـة خـام، عـن طريـق جَمْـع المـواد اللُّغويـة بشـكلها الطبيعـي دون توسـيم. ثـم بـدأ العمـل بشـكل واسـع علـى توسـيم المدونـات في عقـد التسـعينات؛ وذلـك بالتركيـز علـى التحليـل النحـوي وتصنيـف أقسـام الـكلام، واتسـع نطـاق العمـل تدريجيًّـا ليشـمل مختلـف المسـتويات اللُّغويـة وشـبه اللُّغويـة، وتطـوَّر مفهـوم المدونـات اللُّغويـة ليشـمل المـواد اللُّغويـة متعـددة الوسـائط (النص – الصوت – الفيديو).

ولعـل أبـرز الجهـود العالميـة لوضـع معايـير للمدونـات وتوسـيمها مشـروع معيـار ترميـز المدونـات (Corpus Encoding Standard – CES) الـذي قدَّمـه *Ide (١٩٩٨)*، ضمـن مبـادرة ترميـز النصـوص المعروفـة باسـم **The Text Encoding Initiative (TEI)**. طُبِّـق هـذا المعيـار

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

لترميز المدونات اللغوية، حيث يهدف إلى تقديم إطار مُوحَّد وشامل لترميز المدونات اللُّغوية بطرق تتيح سهولة التبادل والاستخدام بين الباحثين والمُطوِّرين في مجال معالجة اللُّغة الطبيعية.

تمَّ تطوير CES بالتعاون مع مشاريع أوروبية وأمريكية مثل MULTEXT وEAGLES، وبالتنسيق مع مبادرة ترميز النصوص (Text Encoding Initiative – TEI)، التي تُعَدُّ من أهم الجهود في وَضْع معايير لترميز النصوص الإلكترونية. وتُوفّر TEI مجموعةً من الإرشادات والمبادئ التوجيهية لترميز مختلف أنواع النصوص، بما في ذلك النصوص الأدبية والتاريخية واللُّغوية؛ بهدف تسهيل تحليلها ومعالجتها باستخدام الحاسوب.

ويُركّز معيار CES على تقديم أدلة إرشادية لترميز وتخزين المدونات اللُّغوية، مع الالتزام بمبادئ TEI، ولكنه يُضيف معايير تلائم احتياجات معالجة اللُّغة الطبيعية واللسانيات الحاسوبية. ويشمل ذلك تحديد مستويات مختلفة من الترميز تُناسب الأغراض المتنوعة، بَدءًا من الترميز الأساسي للبنية العامة للنص وصولًا إلى الترميز المتقدم الذي يشمل الوسوم اللُّغوية والتحليلات النحوية.

ومن الجدير بالذكر أن هذه المعايير تغطي كلًّا من المدونات الخام والمدونات الموسمة، حيث تُقدِّم إطارًا منهجيًّا لجَمْع النصوص اللُّغوية وتنظيمها وتخزينها بشكل يُسهِّل استخدامَها في تطبيقات متعددة. وقد خُصِّص لهذين النوعين من المدونات فصلان مستقلان في هذه الوثيقة؛ نظرًا لأهمية كلٍّ منهما ودوره في تطوير نماذج معالجة اللُّغة الطبيعية.

ومن أهم الدراسات التي تطرَّقت إلى معايير تصميم المدونات اللُّغوية ما جاء في دراسة (معايير تصميم المدونة Corpus Design Criteria) لـAtkins وآخرين (١٩٩٣). وتُعَدُّ هذه الدراسة من الأعمال الرائدة في مجال تصميم المدونات، حيث قدَّمت إطارًا نظريًّا ومنهجيًّا لتطوير مدونات لغوية تلبي احتياجات البحث اللُّغوي والمعالجة الحاسوبية للغة. وركَّزت الدراسة على مراحل بناء المدونات، بدءًا من التخطيط الذي يشمل

التصميـم اللُّغـوي وتقديـر التكاليـف والإدارة، مـرورًا بالحصـول علـى الأذونـات القانونيـة اللازمـة مـن أصحـاب حقـوق النشـر، ووصـولًا إلى جَمْـع البيانـات ومعالجـة النصـوص، وسـلطت الضـوء علـى أهميـة تحديـد مجتمـع النصـوص وعمليـة أخـذ العينـات، مـع التأكيـد علـى ضـرورة وصـف العينـة المسـتهدفة بدقـة؛ لضمـان تمثيليـة المدونـة وجودتهـا.

إضافـةً إلى ذلـك، ناقشـت الدراسـة التحديات المرتبطة بترميز النصوص ووَضْـع العلامات (Markup)، سـواء في تحويـل المـواد المكتوبـة أو في نسـخ الـكلام المنطـوق، مُشـيرةً إلى المـيزات الـتي ينبغـي تضمينهـا في الترميـز لـكلٍّ مـن النصـوص المكتوبـة والمنطوقـة. كمـا اسـتعرضت التصنيفـات المختلفـة للمدونـات والنصـوص، وأكَّـدت أهميـةَ وَضْـع معاييـر مشـتركة لتسـهيل تبـادل البيانـات، وتعزيـز التعـاون بـين الباحثـين في مختلـف المواقـع والتخصُّصـات.

وتُعَـدُّ الوثيقـة الـتي حرَّرهـا Martin Wynne بعنـوان *"Developing Linguistic Corpora: a Guide to Good Practice"* (بنـاء المدونـات اللُّغويـة: دليـل إلى الممارسـات الجيـدة) (٢٠٠٥) أحـد المراجـع المهمـة في توجيـه الباحثـين والمُطوِّريـن نحـو بنـاء مدونـات لغويـة مسـتندة إلى معاييـر علميـة وتقنيـة راسـخة، دون التركيـز علـى لغـة بذاتهـا. وتعـرض هـذه الوثيقـة عـدة فصـول كتبهـا باحثـون متخصصـون في اللسـانيات الحاسـوبية، وقـد ركَّـزت علـى مراحـل جوهريـة مـن عمليـة بنـاء المدونـة، **أهمهـا:**

- **مبـادئ بنـاء المدونـة:** آليـات انتقـاء النصـوص وعينـات اللُّغـة، ومـا يجـب أَخْـذه في الاعتبـار عنـد تجهيـز مدونـة لأغـراض بحثيـة مختلفـة. وتعـرض الوثيقـة الأسـس العامـة لبنـاء المدونـات، بـدءًا مـن تحديـد الجهـة أو الأفـراد المسـؤولين عـن إنشـاء المدونـة، والأهـداف المرجـوَّة منهـا، وصـولًا إلى كيفيـة أَخْـذ عينـة لغويـة تضمـن تمثيليـة المدونـة وتوازنهـا. كمـا تناقـش الدراسـة حجـم المدونـة ونطـاق تغطيتهـا، ودور ذلـك في تحقيـق الدقـة والشـمول، مـع التركيـز علـى المعاييـر الـتي تسـاعد في بنـاء مدونـة قـادرة علـى تقديـم صـورة وافيـة عـن اللُّغـة أو اللهجـة المُسـتهدَفة.

- **إضافة الوسوم اللُّغوية:** حيث تستعرض الوثيقة أنظمة التوسيم اللُّغوي، وأهمية الالتزام بالمعايير المتعارف عليها لتحقيق قابلية المشاركة وإعادة الاستخدام. وهذا الجزء ينطبق على وثيقة "معايير توسيم البيانات اللُّغوية".
- **البيانات الوصفية (Metadata):** تُوضِّح الوثيقة البيانات الوصفية (Metadata) التي ترافق المدونة وطرق تمثيلها وتنظيمها؛ لتسهيل الوصول والفهم. وتشدد الوثيقة على ضرورة جمع هذه البيانات وتنظيمها وفصلها عن النص الأصلي، فضلًا عن توضيح كيفية استخدام تنسيقات معيارية مثل XML أو JSON، وتحديد ما يلزم تضمينه من معلوماتٍ رئيسة حول المصدر والتاريخ والبيئة اللُّغوية للمدونة، بما يُسهِّل توثيقها واستعادتها ودراستها في وقت لاحق.
- **ترميز الحروف (Character encoding):** حيث تُوضِّح ترميز الحروف بوصفه عنصرًا جوهريًّا في بناء المدونات، ولا سيما عند التعامل مع لغات تستخدم أنظمة كتابة مركّبة مثل العربية. ويُوصي المُؤلِّف بضرورة اعتماد ترميز مُوحَّد (كـ UTF-٨) لتمكين دعم اللُّغات المتعددة، ومنها العربية؛ وذلك لضمان التوافق مع أدوات المعالجة المختلفة، والحفاظ على الاتساق في قراءة النصوص عبر الأنظمة والمنصات المتنوعة.
- **بناء المدونات الصوتية:** حيث تناقش الوثيقة الجوانب المنهجية والفنية المتعلقة ببناء مدونات اللُّغة المنطوقة (المحكية)، بما يشمل كيفية جَمْع التسجيلات الصوتية أو المحادثات، وطرق تفريغ محتواها، ومعايير جودة البيانات المنطوقة. كما تسلط الدراسة الضوء على التحديات التي يواجهها الباحثون في جمع هذه البيانات وتفريغها وترميزها، عند التعامل مع اللغة المنطوقة.
- **الأرشفة والتوزيع والصيانة:** تتطرق الوثيقة إلى توزيع المدونات وتوفيرها للباحثين وحفظها وإجراءات حمايتها من التلف الرقمي على المدى البعيد، مع مراعاة جوانب قانونية وتقنية مختلفة؛ بما يضمن استمرارية وصول المدونة للأجيال اللاحقة من الباحثين.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً



وتـبرز أهميـة هـذه الوثيقـة في كونهـا تُوفِّـر إرشـاداتٍ عمليـةً وتفصيليـةً للتعامـل مـع تحديـات بنـاء المدونـات بمسـتويات متعـددة (نصيـة، صوتيـة، مرئيـة). كمـا تعكـس اهتمامًـا بتوثيـق عمليـة الجمـع والتوسـيم والترميـز وفقًـا لأفضـل الممارسـات التقنيـة والقانونيـة؛ ممـا يجعـل هـذه التوصيـات قابلـة للتطبيـق في البيئـات البحثيـة الحديثـة.

ومـن الدراسـات العربيـة الرائـدة في مجـال بنـاء المدونـات اللُّغويـة يـأتي كتـاب "المدونـات اللُّغويـة العربيـة: بناؤهـا وطرائـق الإفـادة منهـا" (العصيـمي، ٢٠١٥)، الـذي يسـتعرض في خمسـة مباحـث عـددًا مـن الموضوعـات الـتي تتصـل بالمدونـات اللُّغويـة العربيـة وسـبل الاسـتفادة منهـا. تسـتهل الدراسـة بالمبحـث الأوّل: **"المدونـات اللُّغويـة وكيفيـة الإفـادة منهـا"**، مُسـتعرضةً تعريـف المدونـة اللُّغويـة، وطـرق بنائهـا، والغايـات البحثيـة والتطبيقيـة الـتي يمكـن تحقيقهـا عـن طريـق توظيفهـا. ثـم تتنـاول في المبحـث الثـاني **(مدونـات المتعلمين)**، حيـث تُسـلِّط الضـوءَ على أهميـة بنـاء مدونـات خاصـة بالمتعلمـين (مثـل دارسي العربيـة لغـةً ثانيـةً)، وآليـات تصنيـف تلـك المدونـات وتوسـيمها، مـع رصـد أوجـه الاسـتفادة مـن الأخطـاء أو الظواهـر اللُّغويـة عنـد هـذه الفئـة مـن الناطقـين بغـير العربيـة.

وفي المبحـث الثالـث **(تصميـم المدونـات اللُّغويـة وبناؤهـا)**، جـاء التركـيز على الجانـب الإجرائي لعمليـة بنـاء المدونـات، حيـث أشـار المُؤلِّـف إلى المعايـير الرئيسـة في اختيـار النصـوص وعينـات اللُّغـة، والحجـم المناسـب للمدونـة، وآليـات تحقيـق التـوازن والتمثيـل فيهـا. وكيفيـة ضمـان توازنهـا وتمثيليتهـا.[1] ويناقـش هـذا المبحـث -أيضًـا- ضوابـط أخـرى تتعلَّـق بحقـوق الملكيـة الفكريـة في جَمْـع النصـوص وأدوات التوثيـق والتنظيـم.

---

(١) **تـوازن المدونـة** يعـني: **أن تحتـوي على نصـوص أو بيانـات لغويـة موزّعـة بشـكل عـادل ومـدروس على مختلـف الأنـواع أو الفئـات اللغويـة الـتي تمثّـل الظاهـرة المدروسـة.** إذا كنـت تنشئ مدونـة لغويـة للغـة العربيـة مثـلًا، فيُتوقـع أن تـوازن بـين:

- **الأنواع النصية:** مقالات، روايات، تغريدات، خطب، محادثات يومية...
- **الأساليب:** لغة فصيحة، عامية، لغة الصحافة، لغة الأطفال...
- **الجهات الجغرافية:** نصوص من مختلف الدول أو المناطق الناطقة بالعربية.
- **الأنماط الاجتماعية:** فئات عمرية أو تعليمية مختلفة.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

أمّا المبحث الرابع "لسانيات المدونات: نماذج وتطبيقات في لغة الصحافة العربية"، فيستعرض تطبيقات عملية للمدونات في دراسة أنماط اللُّغة العربية الحديثة وتحليل النصوص الصحفية، مُتطرقًا إلى بعض الظواهر اللُّغوية والأسلوبية التي يمكن استخلاصها من كميات كبيرة من نصوص الصحف، ودور ذلك في الأبحاث اللُّغوية الحاسوبية.

ويأتي المبحث الخامس المعنون بـ "البحث اللُّغوي في المدونات العربية الحاسوبية بين المُمكن والمحتمل والمأمول"، ليُوضِّح أفقَ التطوير والابتكار في مجال المدونات العربية، فيناقش الرهانات والمُعوِّقات المحتملة، والفرص الواعدة التي تنتظر الباحثين في ضوء التطوُّرات السريعة في أدوات الحوسبة والذكاء الاصطناعي.

وعلى الرغم مما تتَّسم به هذه الدراسة من تنوُّع وثراء، لا تزال الحاجة ماسّة إلى وثائق مرجعية أكثر تخصُّصًا، تُحدِّد المعايير الواجب اتباعها عند بناء المدونات اللغوية العربية، ولا سيما تلك التي تشكِّل حجر الأساس في معالجة اللغة الطبيعية وتطبيقاتها. ومن هنا تنطلق الوثيقة الراهنة، ساعيةً إلى تجميع أبرز المعايير المعتمدة عالميًّا، وإعادة صياغتها بما يراعي الخصوصيات اللغوية والثقافية للعربية، على نحوٍ يسهم في ردم الفجوة المعرفية، وتعزيز تبادل البيانات بين الباحثين، والارتقاء بجودة النماذج الحاسوبية المعتمدة على اللغة العربية.

واستعرض Zeroual (٢٠١٩) في دراسته – تحت عنوان: **بناء المدونات العربية: المفاهيم، والمنهجيات، والأدوات، والتجارب** – المراحلَ التاريخية لبناء المدونات العربية، مُوضِّحًا أن المدونات قبل عقد الستينات (١٩٦٠) كانت تُسمَّى "ما قبل المدونات

---

**والمقصود بـ "تمثيلها": قدرة عيناتها المُمثِّلة للاستعمال اللُّغوي، أو قدرتها على تمثيل التنوع اللغوي الموجود في المجتمع أو في السياق الذي أُنشئت المدونة من أجله**، سواء من حيث:

- **الأنواع النصية:** مثل المحادثات، المقالات، الأخبار، الروايات...
- **الأساليب:** رسمية أو غير رسمية
- **المستويات اللغوية:** لهجات، فصحى، تخصصية...
- **القطاعات الاجتماعية:** الفئات العمرية، التعليمية، الجغرافية...

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً



الإلكترونيـة" (pre-electronic) حيـث كانـت تعتمـد الطـرق التقليديـة في توثيـق اللُّغـة عـن طريـق الكتابـة باليـد، مُضيفًـا أن بنـاء واسـتخدام المدونـات ليـس اختراعًـا جديـدًا بالكامـل؛ بـل يتمثَّـل الجانـب الجديـد فيهمـا في توطـين مصطلـح "مدونـة" (corpus)، ووضـع معايـير منهجيـة لإنشـائها، إلى جانـب توظيـف التقنيـات الحديثـة في جمـع البيانـات اللغويـة، وتنظيمهـا، واسـتعراضها، والبحـث فيهـا. وأعـاد المُؤلِّـف جـذور بنـاء المدونـات إلى العالـم العـربي (الخليـل بـن أحمـد الفراهيـدي) مستشـهدًا بأعمالـه في القـرن الثامـن الميـلادي في جمـع نصـوص عـدة لبنـاء معجمـه العـين، الـذي يُعَـدُّ مـن أوائـل الأعمـال المعجميـة في تاريـخ اللُّغـات البشـرية.

**ومن أهم أمثلة المدونات العربية المتاحة:**

١- مدونـة مسـمَّاة بـ"عشـرة عشـرة" (ArTenTen (Arts et al., 2014: وهـي مدونـة عربيـة مـن عائلـة مدونـات TenTen أو "عشـرة عشـرة" (Jakubíček et al., 2013)، وتضـمُّ أكـثر مـن ٥٠ لغـةً مختلفـةً، مسـتخلصة مـن شـبكة الإنترنـت. وتحتـوي نسـختها العربيـة علـى مـا يربـو عـن خمسـة مليـارات كلمـة.

٢- المدونـة العربيـة لمجمـع الملـك سـلمان (مدونـة مدينـة الملـك عبـد العزيـز للعلـوم والتقنيـة أو "المدونـة العربيـة" سـابقًا) (Al-Thubaity, 2015): مدونـة لغويـة عربيـة تضـمُّ أكـثر مـن مليـار كلمـة ممـا دُوِّن باللُّغـة العربيـة ابتـداءً مـن عصـر مـا قبـل الإسـلام وحتـى العصـر الحديـث، وتشـمل نصوصًـا مـن بلـدان مختلفـة، مـع الأخـذ في الحسـبان طبيعـة وحجـم النشـاط الفكـري لكل فـترة. وتتنـوَّع أوعيـة النشـر فيهـا؛ مثـل: المخطوطـات، والصحـف، والكتـب، والمجـلات، والدوريـات العلميـة، ونحوهـا، والسـائد مـن المجـالات العلميـة والفكريـة المختلفـة، مثـل: المعتقـدات، وعلـوم العربيـة، والعلـوم الطبيعيـة، والفنـون، وغيرهـا.

٣- مدونـة المئـة مليـار كلمـة عربيـة: وهـي مدونـة نصيـة مصدرهـا (Aloui et al., 2024)، وهـي مجموعـة مـن النصـوص العربيـة المُسـتخرَجة مـن مخـزن common crawl، الـذي يتضمَّـن مجموعـة مـن نصـوص الشـبكة المُحدَّثـة دوريًّـا، ويتجـاوز عـدد

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

كلماتها ١٠١ مليار كلمة، ويصل حجمها إلى ٤٠٠ جيجا بايت تقريبًا. ويُميز المدونة اعتماد باحثيها على منهج متكامل من مرحلة جَمْع بيانات المدونة ثم تنقيتها بأدوات المعالجة الحاسوبية إلى مرحلة الحفظ والتخزين والنشر. وقد تم تنزيل مجموعة جزئية من المخزن العملاق من الأسبوع ٣٩ سنة ٢٠٢١م إلى أسبوع ٢٧ سنة ٢٠٢٢م، وفيها نحو ٢٥٠ موقعًا عربيًّا، ثم عُولِجت النصوص بحذف القصير منها، وإزالة المكتوب غير المرغوب فيه كالروابط وترميزات لغة التوصيف (HTML)، ثم حُفظت كل قطعة نصية مع تاريخها ورمز تعريفي مُميَّز في صيغة ملف json.. ومما يعيب المدونة أن عددًا من نصوصها ما زال يحتوي على مُهملات نصية قد تضر بالنماذج المُستعملة لها، مثل كثرة النصوص المكررة والكلمات الغريبة.

٤- مدونة المقالات العربية -خمس مئة ألف مقالة- (Altamimi & Alyaba 2024): وهي مدونة نصية جُمِعَت من عدد من الجرائد الإخبارية، وتتضمَّن ٥٠٠ ألف مقالة عربية مُوَسَّمة بحسب الفئات التالية: (ترفيهي، سياحي، صحي، سيارات، فن، محلي، تقني، اقتصادي، سياسي، رياضي) مرتبة من أقل النصوص إلى أكثرها عددًا. وجُمعت المدونة باستعمال عدد من أدوات استخراج النصوص اللُّغوية. وشمل تاريخ المقالات المسحوبة الفترة الزمنية من ٢٠٢١/١/١م إلى ٢٠٢١/١٢/٣١م. ويمكن استعمال بيانات المدونة في اختبارات المعالجة الآلية للغة، والتوليد والتصنيف، وقياس نسبة التشابه الدلالي، وغيرها.

٥- مدونة أبي الخير للنصوص الإخبارية، المُكوَّنة من مليار ونصف مليار كلمة (١،٥ مليار كلمة) (Abu El-khair, n.d.)، وتحتوي المدونة على نصوص مجموعة من عشر صحف عربية من مختلف الدول العربية (السعودية، والإمارات، وسوريا، واليمن، ومصر، والجزائر، والكويت، ولبنان)، وتشمل أكثر من ٥،٥ ملايين مقالة، كُتبت في فترة زمنية تعادل ١٤ سنة تقريبًا.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً



والجديـر بالذكـر أنّ مُطـوِّري المدونـة اسـتعملوا عـددًا مـن أدوات اسـتخراج النصـوص، ولا سـيما أداة Offline Explorer Pro الـتي تمـيزت بالسـرعة وثبـات الأداء مقارنـةً بـالأدوات الأخـرى.

٦- مدونـة ملخصـات المقـالات العربيـة (Alhamadani, 2023): مدونـة نصيـة تحـوي أكـثرمـن ثمانيـة ملايـين مقالـة مرفقـة بملخّصاتهـا، وقـد جُمعـت مـن محتـوى الويـب العـربي، لتغطـي مختلـف الـدول العربيـة خـلال الفـترة الممتـدة مـن ١٩٩٩ إلى ٢٠١٩. وتُعَـدُّ هـذه المدونـة مـوردًا غنيًّـا لدراسـات تلخيـص النصـوص ومعالجـة اللغـة الطبيعيـة.

ومـن أبـرز التحديـات المرتبطـة بهـا قِـدم نصوصهـا، إذ تعـود بعـض المقـالات إلى مـا يزيـد علـى ثلاثـة عقـود، ممـا قـد يؤثـر علـى حداثـة المحتـوى وملاءمتـه لتدريـب نمـاذج اللغـة المعاصـرة.

٧- مدونـة مجمـع الملـك سـلمان العالـمي للغـة العربيـة[1]: وهـي مدونـة لغويـة مكتوبـة غـيرموسَّـمة، تتنـوع مجالاتهـا لتشـمل المجـالات: الدينيـة، والثقافيـة، والاجتماعيـة، والسياسـية، والعلميـة، وغيرهـا، ويـندرج تحـت كل مجـال عـدد مـن الموضوعـات. وتمثـل المدونـة أوعيـة نشـر متعـددة، وهـي: الكتـب، والمجـلات، والرسـائل الجامعيـة، ومواقـع الإنترنـت، والصحـف. وتهـدف إلى إثـراء مصـادر البيانـات اللُّغويـة المُصنَّفـة وتنويعهـا. ويشـير (الجـدول ١) إلى قائمـة تتضمـن تفاصيـل أهـمِّ المدونـات اللُّغويـة.

(1) https://falak.ksaa.gov.sa/corpora/

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

الجدول (١): قائمة تتضمن تفاصيل أهم المدونات اللُّغوية العربية

| اسم المدونة | حجمها | عدد الكلمات | سنة النشر |
|---|---|---|---|
| مدونة تن تن arTenTen العربية | - | ٦,٥ مليارات | ٢٠١٤ |
| المدونة العربية | - | مليار كلمة | ٢٠١٥ |
| مدونة المئة مليار | ٤٠٠ جيجا | ١٠١ مليار كلمة | ٢٠٢٤ |
| مدونة المقالات العربية | - | ٥٠٠ ألف مقالة | ٢٠٢٣ |
| مدونة المقالات وملخصاتها | ٣٠ جيجا | ٨ ملايين مقالة | ٢٠٢٣ |
| مدونة مجمع الملك سلمان العالمي للغة العربية | - | مليون كلمة | ٢٠٢٤ |

ويُلحظ تباين مناهج جمع البيانات في المدونات السابقة، حيث تعتمد غالبيتها على استخلاص النصوص من مواقع الصحف العربية، وهي الطريقة الأكثر شيوعًا. في المقابل، تلجأ بعض المدونات إلى استخدام برامج متخصصة في جمع النصوص من الويب، في حين تُعَدُّ المدونة الأولى حالة متميّزة، إذ جُمعت بياناتها من Common Crawl، بعد تصفيتها بطريقة مبتكرة وفق شروط محددة، إلا أن مطوّريها لم يشيروا إلى اعتمادهم على منهجية واضحة أو دليل مرجعي في عملية الجمع والبناء. أما مدونة النص العربي ٢٠٢٢، فهي لا تمثّل نتاجًا أصيلًا، بل عبارة عن تجميع لعدد من المدونات النصية المنشورة مسبقًا (ذُكرت القائمة في الورقة البحثية). ويلحظ -كذلك- أن أغلب هذه المدونات تفتقر إلى إطار منهجي دقيق أو معايير محددة لبنائها، فضلًا عن احتوائها على نسبة عالية من البيانات المشوشة التي قد تضعف فاعليتها في مهمات المعالجة الحاسوبية للغة.

ونظرًا إلى أن المدونات اللغوية تُشكِّل الأساس الذي تُبنَى عليه نماذج المعالجة الطبيعية للغة، فإن ازدياد الاعتماد على تطبيقات الذكاء الاصطناعي ومعالجة اللغة الطبيعية يفرض حاجة مُلحة إلى وضع معايير مُوحَّدة وواضحة لبناء هذه المدونات، بما يضمن جودتها وموثوقيتها في دعم التطوير اللغوي الحاسوبي.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً



ويُشار إلى أن الدراسات والممارسات السابقة -سواء المُؤلَّفات في وصف المدونات وطرق بنائها بشكل عام، أو المنشورات الخاصة بمدونات بذاتها- لا تُوضِّح بشكل كافٍ المعايير التي ينبغي استخدامها عند بناء المدونات اللُّغوية، خاصة في اللُّغة العربية. وعلى ذلك، تجمع هذه الوثيقة المعايير العالمية المعتمدة في بناء المدونات اللُّغوية مع التركيز على المعايير الخاصة باللُّغة العربية. وتتضمَّن وثيقة المعايير الحالية عددًا من المعايير العامة لبناء المدونات اللُّغوية العربية التي يندرج تحتها مجموعة من المعايير العامة والفرعية المُقسَّمة وفقًا لمختلف مراحل بناء المدونات اللُّغوية، **وتشمل فئات المعايير الموضوعات الرئيسة الآتية:**

١- مواصفات المدونة.

٢- جَمْع مواد المدونة.

٣- معالجة المدونة.

٤- تخزين المدونة وتوثيق بنائها.

## نطاق المعايير وحدودها:

تُركِّز هذه الوثيقة على ذكر المعايير الخاصة ببناء المدونة اللُّغوية مرتبةً حسب تسلسلها المنطقي في مراحل عملية بناء المدونة اللُّغوية، بدايةً من تحديد معايير ومواصفات المدونة، وجَمْع المواد اللُّغوية أثناء عملية البناء، ومن ثَمَّ معالجة المدونة، وأخيرًا التركيز على سرد معايير تخزين المدونة وتوثيق بنائها. ولا تتطرق هذه الوثيقة إلى مرحلة توسيم المدونة التي تعتبر عملية مستقلة عن جَمْع المدونات الخام وبنائها.

## قائمة معايير بناء المدونات اللُّغوية[1]:

| المجال | المعيار العام | المعيار التفصيلي | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| **أولًا: تحديد مواصفات المدونة** | ١. تحديد الهدف من المدونة | أ. وضوح الهدف المُحدَّد من بناء المدونة. | تحديد الهدف من جَمْع المدونة بشكل واضح، مثل: دراسة ظاهرة لغوية معينة. | وجود وثيقة مكتوبة توضّح الهدف | مراجعة الوثائق المصاحبة للمدونة |
| | | ب. وجود غرض أو أغراض مُحدَّدة من بناء المدونة. | تحديد الغرض (أو الأغراض) من بناء المدونة، مثل: أن تسهم في فَهْم الظاهرة اللُّغوية المُستهدَفة على نطاق معين. | وجود الأغراض المحددة في وثيقة المدونة | تحليل محتوى وثيقة توصيف المشروع |
| | | ج. مطابقة الأهداف مع المخرجات. | مراجعة الأهداف لبناء المدونة بعد تحديد مواصفاتها؛ للتحقُّق من أن مخرجات المواصفات تطابق الأهداف. | نسبة التوافق بين الأهداف والمخرجات الفعلية | مقارنة تحليلية بين أهداف الوثيقة والبيانات المنتجة (المخرجات). |

(١) تم إدراج مؤشرات أداء متوسطة، لا تتطلب - غالبا - إجراءات مطولة، بحيث يمكن الإفادة منها بيسر وسهولة.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعيار التفصيلي | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | ٢. تحديد النوع اللُّغوي للمدونة | أ. تحديد النوع أو الأنواع اللُّغوية المُستهدَفة في المدونة. | تحديد الأنواع أو النوع اللُّغوي للمدونة، مثل: التسجيلات الصوتية للغة المحكية، أو النصوص المكتوبة. | وضوح تصنيف النوع اللغوي في الوثائق | مراجعة وصف المدونة الفني |
| | | ب. تحديد النوع من حيث العموم والخصوص. | تحديد النوع من حيث العموم والخصوص، مثل: المدونات الخاصة لغير المتحدثين باللُّغة. | وجود توصيف صريح لمدى خصوصية المدونة | فحص وثائق توصيف المدونة |
| | ٣. تحديد الوعاء اللُّغوي | أ. توضيح الوعاء اللُّغوي. | توضيح الوعاء اللُّغوي للمدونة، مثل: الكتب، والمقالات، ووسائل التواصل الاجتماعي. | وجود بيان تفصيلي للوعاء اللغوي | تحليل محتوى توصيف المصدر |
| | | ب. ذِكْر عدد الأوعية اللُّغوية. | يحدّد عدد الأوعية اللغوية في حال وجود أكثر من وعاء لغوي. | عدد الأوعية اللغوية المختلفة المحددة | فحص التصنيفات داخل توصيف المحتوى |
| | ٤. تحديد النطاق الجغرافي | أ. تحديد النطاق الجغرافي الذي تُغطّيه المدونة. | تحديد المناطق الجغرافية التي تُغطيها المدونة؛ لضمان تمثيل لهجات ومناطق مختلفة. | وجود النطاق المحدد جغرافيًا | تحليل التوزيع الجغرافي للبيانات |
| | | ب. توضيح عدد المناطق التي تغطيها المدونة. | تحديد عدد المناطق التي تغطيها المدونة. | وجود المناطق الممثلة في المدونة | فحص بيانات المتغير الجغرافي في البيانات الوصفية |
| | ٥. تحديد الفترة الزمنية | أ. تحديد بداية الفترة الزمنية ونهايتها للمدونة. | تُحدَّد بداية الفترة الزمنية ونهايتها للمدونة. | وجود توثيق زمني واضح للمدونة | مراجعة البيانات الوصفية أو وثيقة المشروع |

| المجال | المعيار العام | المعيار التفصيلي | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ب. تحديد عدد الفترات الزمنية ومدتها. | توضيح عدد الفترات الزمنية ومدتها، في حال كانت المدونة تغطي فترات زمنية مختلفة. | عدد الفترات الزمنية المفصّلة | تحليل محتوى التوثيق الزمني |
| | ٦. التحقق من شمول المدونة وتمثيل عيناتها وتوازنها | أ. ذِكْر عدد الاستعمالات اللُّغوية المُغطَّاة. | التحقق من أن المدونة تشمل استعمالات لغوية مختلفة، تغطي الهدف من بناء المدونة. | عدد الاستعمالات اللغوية المشمولة | تحليل التصنيفات اللغوية في المحتوى |
| | | ب. توضيح عدد العينات المُمثِّلة للاستعمال اللُّغوي. | يُحدَّد عدد العينات المُمثِّلة للاستعمال اللُّغوي، مع التحقُّق من تمثيل العينات لهذه الاستعمالات. | نسبة العينات المعرفة بدقة | مراجعة تصميم العينة في الوثائق |
| | | ج. تحديد نسبة كل عينة من حجم المدونة الكلي. | أن تكون المدونة متوازنةً في جمع العينات من أنواع أدبية واستعمالات لغوية مختلفة دون أن يطغى نوع معين على الأنواع الأخرى. | وضوح التوزيع النسبي للعينات | فحص وثيقة التوزيع العيني وتحليل البيانات |
| | ٧. تحديد حجم المدونة | أ. تحديد عدد الكلمات الإجمالي. | يُوضَّح عدد الكلمات الكلي في المدونة كبيانات وصفية لها. ويُحدَّد العدد المُستهدَف التقريبي سلفًا قبل جمع المدونة. | وجود إجمالي عدد الكلمات في المدونة | استخدام أدوات برمجية لحساب عدد الكلمات. |
| | | ب. تحديد عدد الكلمات المُمَيَّزة دون تكرار. | يُوضَّح عدد الكلمات المميزة (بدون تكرار) كبيانات وصفية للمدونة. | عدد الكلمات المميزة في المدونة | أدوات تحليل المحتوى واستخراج المفردات |

| المجال | المعيار العام | المعيار التفصيلي | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ج. تحديد عدد النصوص. | يُوضَّح عدد النصوص المختلفة، مثل: عدد المقالات التي جُمِعت منها المدونة، أو عدد المقاطع الصوتية. ويُحدَّد العدد المُستهدَف التقريبي سلفًا قبل جمع المدونة. | مجموع النصوص المفردة | تحليل قاعدة البيانات النصية |
| | | د. تحديد عدد الساعات بالنسبة للمدونات الصوتية. | يُوضَّح عدد الساعات الإجمالي، وكذلك عدد الساعات لكل مقطع، ويُحدَّد العدد المُستهدَف التقريبي سلفًا قبل جمع المدونة. | عدد الساعات الموثقة في المدونة الصوتية | مراجعة البيانات الوصفية للملفات الصوتية |
| **ثانيًا: معايير جَمْع مواد المدونة** | ١. تحديد المصادر والتحقُّق منها | أ. تحديد مصادر جَمْع المدونة. | تُحدَّد المصادر المُستهدَفة لجَمْع المعلومات منها، بحيث تتناسب مع الهدف، مثل أسماء الصحف في دولة ما. | عدد المصادر الموثقة ضمن وثيقة المدونة | مراجعة وثيقة توصيف المدونة |
| | | ب. التحقُّق من دقة المصار من الناحيتين: اللُّغوية، والثقافية. | التحقُّق من أن المصادر دقيقة لغويًّا، وتتلاءم مع الثقافة المُستهدَفة في المدونة. | نسبة المصادر التي تستوفي معايير الدقة اللغوية والثقافية | تقييم يدوي أو نصف آلي بناء على معايير مسبقة |
| | ٢. موافقة رخصة الاستخدام لمواد المدونة | أ. التحقق من التوافق مع رخصة استخدام المواد اللُّغوية. | رُخَص الاستخدام تختلف فيما بينها من حيث إمكانية إعادة نشر المحتوى والاستفادة منه. | عدد المواد التي رُوجِعت رخصها | مراجعة قانونية وفنية لرخص الاستخدام |

| المجال | المعيار العام | المعيار التفصيلي | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ب. الحصول على الموافقات اللازمة للملكية الفكرية. | في حال تطلّبت أهداف المدونة الحصول على ترخيص لاستخدام المواد اللغوية، سواء من الكاتب أو الناشر، بما يتوافق مع غايات المدونة؛ فإنه ينبغي استخراج هذا الترخيص بشكل رسمي. | عدد الموافقات القانونية المحصّلة | متابعة قانونية وتوثيق رسمي للموافقات |
| | | ج. تحديد رخصة المدونة المناسبة والمتوافقة مع شروط رُخص المواد اللُّغوية. | أن تكون الرخصة المعتمدة في المدونة غير متعارضة مع الرخص الأصلية للمواد اللغوية التي جُمعت منها، إذ إن بعض الرخص تُقيِّد إعادة نشر المحتوى أو استخدامه برخصة مختلفة أو غير متوافقة، مما قد يعرّض المشروع لمشكلات قانونية. | وجود رخصة متوافقة مع كل رخص المواد المجمعة | تحليل قانوني لمدى توافق الترخيص |
| | ٣. تحديد أدوات الجَمْع | أ. تحديد الأدوات المناسبة لجَمْع مواد المدونة. | تحديد الأدوات المناسبة لجَمْع مواد المدونة، مثل زواحف الإنترنت. | عدد الأدوات المحددة المناسبة | مراجعة تقنية لأدوات الجمع مقابل نوعية المادة |
| | | ب. توافق أدوات الجمع مع التنظيمات واللوائح الخاصة بالبيانات. | مراعاة توافق الأدوات مع التنظيمات واللوائح الخاصة بالبيانات، وجمعها وتوثيقها لبيانات كافية حول المصادر الأصلية للنصوص | نسبة الأدوات المتوافقة مع لوائح الحماية | مراجعة توافق الأداة مع المعايير القانونية |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعيار التفصيلي | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ج. قدرة أدوات الجمع على مراعاة تمثيل فئات متعددة للبيانات اللُّغوية وفقًا للمحتوى اللُّغوي المُستهدَف. | التأكد من تمثيل بيانات المدونة للتنوع اللُّغوي واللهجي والدلالي وفقًا لأغراض بناء المدونة. | عدد الفئات اللغوية الممثلة عبر الأداة | تحليل تنوّع ناتج الأداة بناء على التصنيفات اللغوية |
| **ثالثًا: معايير معالجة المدونة** | ١. تنظيف النصوص من الرموز وتوحيد هيكلها | أ. حذف علامات التنسيق الخاصة، مثل: HTML. | عند جَمْع البيانات اللُّغوية من مصادر مثل الإنترنت، قد يتخللها رموزٌ خاصةٌ، مثل رموز لغة HTML؛ مما يستدعي حذفها لضمان نظافة البيانات وصلاحيتها للمعالجة الحاسوبية. | نسبة النصوص الخالية من الرموز غير اللغوية | فحص آلي للنصوص باستخدام أدوات التنظيف |
| | | ب. توحيد هيكلة البيانات في المدونة. | بعض المصادر مثل الصحف الإلكترونية، قد يكون فيها العنوان ونص المقالة الأساسية بتنسيق مختلف باستخدام رموز خاصة. وعليه، ينبغي معالجة هذه النصوص اللُّغوية بشكل متسق. | نسبة النصوص ذات الهيكلة الموحدة | تحليل هيكلي للنصوص حسب القوالب الموحدة |
| | | ج. إضافة الرموز الخاصة بشكل متسق. | عند الحاجة لإضافة رموز خاصة، مثل: الرموز التي تفرق بين العنوان والنص في داخل المقالة، ينبغي أن تكون الرموز متسقةً. | عدد الرموز المضافة بشكل موحّد | مراجعة يدوية أو آلية لاستخدام الرموز |

| المجال | المعيار العام | المعيار التفصيلي | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | د. الاتساق في مواد المدونة المُعالَجة. | تطبيق جميع المعايير الخاصة بتنظيف النصوص والتنسيق بشكل متسق بين كل مواد المدونة، دون اختلاف بين مادة وأخرى. | نسبة المواد المعالجة المتسقة مع المعايير | مطابقة النتائج مع لائحة معايير التنسيق |
| | ٢. استخدام ترميز معياري | أ. توافق الترميز (encoding) المُستخدَم مع المنصات المختلفة. | تخزين المدونة الموسمة بترميز معياري مثل UTF-٨ وصيغة ملفات معيارية مثل xml بحيث تكون متوافقةً مع مختلف المنصات وأنظمة التشغيل. | نسبة الملفات المتوافقة ترميزيًا | فحص الترميز باستخدام أدوات كشف الترميز |
| | | ب. التحقق من اتساق الترميز المُستخدَم. | اتساق الترميز المُستخدَم في جميع مواد المدونة، وعدم اختلافه. | عدد الملفات ذات ترميز موحد | مراجعة عيّنة آلية من ملفات المدونة |
| **رابعًا: معايير تخزين المدونة وتوثيق بنائها** | ١. ضبط مقروئية المدونة | أ. تحسين درجة المقروئية، | **تخزين المدونة بشكل ييسر للمستخدم تصفحها وليس فقط للآلة،** | مدى وضوح النص وسهولة قراءته وفهمه. | اختبار زمني لتجربة الوصول. مثل معدل إتمام قراءة النص. |
| | | ب. تصميم واجهات مُستخدِم مناسبة للمدونة. | وجود أدوات مثل واجهة المُستخدِم الرسومية ييسّر التعامل مع المدونة؛ وذلك عن طريق تقسيمها بشكل منطقي يتناسب مع الهدف منها؛ كأن تُقسَّم إلى فترات زمنية مختلفة أو تُرتَّب بحسب الموضوعات. | نسبة الواجهات المطابقة لمعايير تجربة المستخدم | مراجعة تصميمية باستخدام قائمة تحقق UX |
| | | ج. تسجيل ملاحظات المستخدمين ومتابعتها. | وجود آلية لتسجيل ملاحظات المستخدمين حول استخدام المدونة. | عدد الملاحظات المسجلة والمفعّلة | نظام متابعة واستطلاع دوري |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعيار التفصيلي | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | ٢. تحسين عمليات الوصول والبحث | أ. تحسين درجة الوصولية لدرجة عالية | التحقق من سهولة الوصول إلى المدونة واستخدامها؛ وذلك بقياس عدد الروابط التي يُتطلب الدخول لها للوصول لمعلومة معينة. | نسبة النجاح في الوصول من أول محاولة | تحليل سجلات الوصول |
| | | ب. توافر واجهات متعددة وموثوقة للوصول إلى بيانات المدونة. | استعمال طرق متعددة للوصول إلى بيانات المدونة، مثل:<br>- الواجهة الرسومية، Graphical User Interface (GUI).<br>- مُوجّه الأوامر Command-Line Interface -CLI<br>- واجهة برمجية للتطبيقات Application Programming Interface )API | عدد أنواع الواجهات المتاحة | مراجعة فنية للمنصة |
| | | ج. تحسين طرق البحث المتعددة في المدونة وفقًا لسمات النصوص المختلفة. | استعمال طرق متعددة لفهرسة بيانات المدونة، والبحث في محتواها. | عدد آليات البحث النشطة | اختبار عملي لآليات البحث |
| | ٣. فصل البيانات الوصفية عن مواد المدونة | أ. تحسين كفاءة فصل البيانات الوصفية. | أن تكون البيانات الوصفية [١] مفصولةً عن المواد الرئيسة، وتشمل البيانات الوصفية تاريخ جمع المادة. وتُفصَل عن المواد الأصلية بحيث لا تتداخل معها. | نسبة الملفات المفصولة بشكل صحيح | تحليل آلي لبنية البيانات |

(١) البيانات الوصفية :(Metadata) هي معلومات تصف وتُعرِّف وتُنظِّم البيانات الأساسية، دون أن تكون جزءًا منها، مثل: تاريخ جمعها، واسم المؤلف، والعنوان، والكلمات المفتاحية. كما تتضمن نوع المحتوى وصيغته، واللغة المستخدمة، والجهة الناشرة أو المصدر. يُراعى فصل هذه البيانات عن المادة الأصلية لتفادي التداخل معها، مع الحفاظ على ترابطها البنيوي بها. وتُسهم هذه البيانات في تسهيل الفهرسة والبحث والتوثيق وضمان جودة التنظيم.

| المجال | المعيار العام | المعيار التفصيلي | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | | ب. التحقق من سلامة البيانات من الخلل عند فصل البيانات الوصفية عن المواد الأساسية. | التحقُّق من أن عملية الفصل لا تُسبِّب تغيرًا في البيانات الأصلية. | عدد الملفات الخالية من الأخطاء بعد الفصل | اختبار تقني لسلامة الفصل |
| | ٤. عمل نُسخ احتياطية بشكل دوري | أ. عمل نُسخ احتياطية بشكل دوري. | تُنسخ المدونة بشكل دوري كنسخة احتياطية؛ للحفاظ على سلامة البيانات. | عدد النسخ الاحتياطية المنجزة في المدة المحددة | مراجعة جداول النسخ الاحتياطي |
| | | ب. التحقق من استعادة البيانات من النسخ الاحتياطية بشكل سليم. | التحقُّق من أن النسخ الاحتياطية تحافظ على البيانات الأصلية دون خلل. | نسبة النسخ القابلة للاستعادة بنجاح | تجربة استرجاع عينة نسخ |
| | ٥. تطبيق آليات مناسبة للتفريق بين الإصدارات المختلفة | أ. ترقيم الإصدارات المختلفة للمدونة. | تطبيق آلية لتسلسل الإصدارات المختلفة للمدونة. | نسبة الإصدارات المرقمة | تحليل بنية أسماء الإصدارات |
| | | ب. توثيق الإصدارات المختلفة للمدونة. | توثيق التعديلات على أي نسخة جديدة للمدونة بشكل واضح، مثل: إضافة محتوى جديد، أو التعديل على محتوى سابق. | نسبة التغييرات الموثقة | مقارنة بين النسخ وتحليل السجلات |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعيار التفصيلي | الوصف | مؤشر القياس | آلية القياس |
|---|---|---|---|---|---|
| | ٦. المحافظة على أمن بيانات المدونة | أ. تطبيق معايير الأمن السيبراني. | تطبيق معايير الأمن السيبراني لحماية بيانات المدونة من التخريب أو الاختراق أو التعديل غير المصرَّح به؛ وذلك عبر منع الوصول الآلي المُفرط الذي قد يؤدي إلى تعطيلها، إلى جانب فصل نظام التعديل عن نظام القراءة والاطلاع لضمان سلامة المحتوى واستقراره. | عدد معايير الأمان المطبقة | تدقيق أمني حسب قائمة معايير |
| | | ب. حماية بيانات المدونة من الوصول غير المشروع. | تطبيق الطرق التقنية التي تُنظِّم عملية الدخول إلى بيانات المدونة، مثل: استعمال اسم المُستخدِم وكلمة المرور، أو طرق التحقُّق الثنائي للمُستخدِمين. | نسبة النقاط المحمية بنظام دخول آمن | مراجعة نظام الدخول والتحكم |
| | | ج. تطبيق أنظمة حماية البيانات وضمان خصوصيتها. | التأكد من حفظ بيانات المدونة وتخزينها وفقًا لسياسات الخصوصية وحماية البيانات الشخصية. | نسبة التوافق مع قوانين الخصوصية | تدقيق قانوني وفق أنظمة الخصوصية |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

## المراجع :


- العصيـمي، صالـح فهـد؛ وآخـرون. (٢٠١٥). المدونـات اللُّغويـة العربيـة: بناؤهـا وطرائـق الإفـادة منهـا، مركـز الملـك عبـد الله لخدمـة اللُّغـة العربيـة، الريـاض.

- Abu El-khair, I. (2016). *1.5 billion words Arabic Corpus. arXiv*. https://arxiv.org/abs/1611.04033

- Alhamadani, A. (2023). LANS: Large-scale Arabic News Summarization Corpus. In Proceedings of ArabicNLP. *Association for Computational Linguistics.*

- Aloui, M., Chouikhi, H., Chaabane, G., Kchaou, H., & Dhaouadi, C. (2024). 101 Billion Arabic Words Dataset. *arvix*.

- Altamimi, M., & Alyaba, A. (2023). ANAD: Arabic news article dataset ScienceDirect.

- Al-Thubaity, Abdulmohsen O. "A 700M+ Arabic corpus: KACST Arabic corpus design and construction." Language Resources and Evaluation 49 (2015): 721751-.

- Atkins, S., Clear, J. & Lstler, N. (1992). Corpus Design Criteria. Literary and linguistic computing, 7(1), 110.1093 .16-/llc/7.1.1

- Arts, T., Belinkov, Y., Habash, N., Kilgarriff, A., & Suchomel, V. (2014). arTenTen: Arabic corpus and word sketches. Journal of King Saud University-Computer and Information Sciences, 26(4), 357371-.

- Jakubíček, M., Kilgarriff, A., Kovář, V., Rychlý, P., & Suchomel, V. (2013, July). The TenTen corpus family. In 7th international corpus linguistics conference CL (pp. 125127-).

- Liu, Y. (2024). Datasets for Large Language Models: A Comprehensive Survey. *arxiv*.

- Meyer, C. F., Grabowski, R., Han, H. Y., Mantzouranis, K., & Moses, S. (2003). The world wide web as linguistic corpus. In Corpus Analysis (pp. 241254-). Brill.

- Nakov, P. (2014). Web as a Corpus: Going Beyond the n-gram. *Information Retrieval: 8th Russian Summer School,* 185–228.
- Wynne, M. (Ed.). (2005). *Developing Linguistic Corpora: a Guide to Good Practice*. Oxford: Oxbow Books. http://ota.ox.ac.uk/documents/creating/dlc/
- Zeroual, I., & Lakhouaja, A. (2018). Data science in light of natural language processing: An overview. *Procedia Computer Science, 127*, 8291-.

# الفصل السادس

## معايير توسيم المدونات اللُّغوية العربية

- ملخص
- مقدمة
- مدخل تعريفي إلى توسيم المدونات اللُّغوية العربية
- مُسوِّغات بناء المعايير لتوسيم المدونات اللُّغوية العربية
- الدراسات السابقة
- نطاق المعايير وحدودها
- قائمة المعايير



هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

## ملخص

يُقصَـد بالتوسـيم إضافـة معلومـات لغويـة رئيسـة إلى المدونـة اللغويـة في صورتهـا الخـام، وهـو مـا يُمـيز المدونـة الموسـمة عـن المدونـة الخـام؛ إذ لا تقتصـر الأولى علـى جمـع المـواد اللغويـة بصيغتهـا الطبيعيـة، بـل تتطلّـب جهـدًا إضافيًّـا يتمثَّـل في إدراج وسـوم لغويـة مُحـدَّدة ومتفـق عليهـا مسـبقًا. وتُعَـدُّ هـذه الوسـوم عناصـر مصطنعـة لا تُسـتخدم بطبيعتهـا في الاسـتعمال اللغـوي، ممـا يجعـل التوسـيم عمليـة تتطلـب معرفـة لغويـة دقيقـة وتعاونًـا جماعيًّـا. ومـن هنـا كان مـن الضـروري التفريـق بـين معايـير بنـاء المدونـة الخـام، مثـل آليـات اختيـار العينـات، ومعايـير التوسـيم كإعـداد الأدلـة وتدريـب الموسِّـمين، مـع الإشـارة إلى وجـود بعـض المعايـير المشـتركة، كتنظيـم وتخزيـن البيانـات. وبعـد أن عـرض الفصـل السـابق معايـير بنـاء المدونـة الخـام، تـأتي هـذه الوثيقـة لتضـع إطـارًا شـاملًا لمعايـير توسـيم المدونـات اللغويـة العربيـة، بمـا يضمـن جودتهـا وموثوقيتهـا، ويجعلهـا أساسًـا متينًـا في مشـروعات المعالجـة الآليـة للغـة العربيـة وتطبيقـات اللسـانيات، مـع مراعـاة أن تكـون هـذه المعايـير مرنـةً وغـير مُقيَّـدة بتقنيـات بعينهـا؛ لتواكـب التطـور التقـني المسـتمر.

## مقدمة

في إطـار التطـوُّر المتسـارع الـذي يشـهده العالـم في مجـال أنظمـة الـذكاء الاصطناعـي ولا سـيما في أنظمـة الـذكاء الاصطناعـي التوليـدي (GenAI) ذات القـدرات المتقدمـة في إنتـاج النصـوص اللُّغويـة المتنوعـة اعتمـادًا علـى النمـاذج اللُّغويـة الضخمـة؛ بـرزت الحاجـة إلى توافـر بيانـات لغويـة رقميـة عاليـة الجـودة؛ إذ تلعـب هـذه البيانـات دورًا محوريًّـا في تدريـب نمـاذج المعالجـة الآليـة، وأدائهـا للغـة وقيـاس جـودة تطبيقاتهـا المتعـددة.

وتُعَـدُّ (المدونـات اللُّغويـة) (Corpora) مـن أبـرز المـوارد اللُّغويـة الرقميـة في ميـدان المعالجـة الآليـة للغـة، إذ تُشـكِّل الأسـاس الـذي تقـوم عليـه النمـاذج اللُّغويـة الضخمـة. حيـث تلعـب جـودة هـذه المدونـات ودقتهـا دورًا جوهريًـا في تحديـد كفـاءة مخرجـات نمـاذج الـذكاء الاصطناعـي اللُّغويـة، فضـلًا عـن أهميتهـا البالغـة في الدراسـات اللسـانية. ويمكـن تصنيـف المدونـات اللُّغويـة وفـق معايـير متعـددة؛ منهـا: الغـرض مـن اسـتخدامها (عامـة أو متخصصـة)، وطبيعـة المحتـوى (مكتوبـة أو منطوقـة)، وكذلـك حسـب البيانـات الوصفيـة المرافقـة لهـا حيـث تُصنَّـف إلى **فئتـين رئيسـيتين**:

- **المدونات الخام:** تتضمَّن جَمْع المواد اللُّغوية بصورتها الطبيعية وتخزينها رقميًا.
- **المدونات الموسمة:** تحتوي على المواد الخام مُضافًا إليها بيانات وصفية ووسوم لغوية متنوعة، تُثري المحتوى وتُيسِّر تحليله واستثماره.

وتتطلـب المدونـات الموسـمة معرفـةً لغويـةً متعمقـةً وتدريبًـا مُتخصِّصًـا، لكنهـا تُقـدِّم اسـتخدامات أوسـع، وتتيـح للنمـاذج اللُّغويـة والخوارزميـات الحاسـوبية اسـتخلاص رؤى لغويـة أعمـق؛ وهـو مـا يُعـزِّز جـودة مخرجـات تطبيقـات الـذكاء الاصطناعـي. ويمكـن القـول باختصـار: إن المدونـة الخـام تُمثِّـل اسـتخدامًا لغويًّـا طبيعيًّـا، في حـين تشـتمل المدونـة الموسـمة علـى معلومـات لغويـة إضافيـة تهـدف لإثـراء وتوضيـح المحتـوى اللُّغـوي الخـام.

وتهـدف هـذه الوثيقـة إلى صياغـة عـدد مـن المعايـير القياسـية الـتي يمكـن الاعتمـاد عليها لضبـط جودة عمليـة توسـيم المدونـة اللُّغويـة العربيـة. وتسـتهل الوثيقـة بعـرض أهـم المصطلحـات المُسـتخدَمة فيهـا، ثـم يتبـع ذلـك مدخـل تعريفـي بمعايـير توسـيم المدونـات اللُّغويـة وحدودهـا، مـع ذكـر أهـم دوافـع وَضْـع هـذه المعايـير، واقـتراح بعـض الطـرق لتطبيقهـا. كمـا تتنـاول أهـم الأدبيـات السـابقة المتعلقـة بتوسـيم المدونـات اللُّغويـة، ثـم تُختتـم بمعايـير توسـيم المدونـات.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

## مدخل تعريفي:

يشير التوسيم اللُّغوي إلى عملية إثراء المدونة الخام -سواء كانت مكتوبة أو منطوقة- عبر إضافة مستويات وطبقات جديدة من البيانات والمعلومات اللُّغوية؛ فتتحول المدونة من نصوص مجردة أو ملفات صوتية، إلى نصوص أو أصوات يرتبط بها كنز من البيانات اللُّغوية المنظمة. ويُشار أحيانًا إلى هذه المستويات الإضافية بالوسوم اللُّغوية، وتشمل خصائص لغوية مختلفة، مثل: أقسام الكلام، وأشجار التحليل النحوي، والتراكيب الاسمية، وغيرها. ويتضمَّن نطاق التوسيم اللُّغوي مستويات لغوية عدة؛ تشمل: المستوى الصوتي، والمستوى الصرفي، والمستوى النحوي، والمستوى الدلالي، والمستوى التداولي. ولكلِّ نوع من هذه الوسوم ميزةٌ إضافيةٌ تُعزِّز من استعمال المدونة في تطبيقات المعالجة الآلية المختلفة. وقد أثبتت الدراسات فوائد جمَّة لاستعمال المدونات الموسمة؛ إذ تُعَدُّ شرطًا رئيسًا لتدريب وتقييم نماذج تعلُّم الآلة والذكاء الاصطناعي في تطبيقات معالجة اللُّغة الطبيعية المختلفة، كما تُستعمل هذه المدونات في الدراسات اللسانية، حيث تساعد الباحثين في اللسانيات التطبيقية وتحليل الخطاب على استكشاف الظواهر اللُّغوية وصياغة النظريات اللُّغوية، ودراسة الاستعمالات اللُّغوية المختلفة وتحليلها. ومن أمثلة التطبيقات اللُّغوية الحاسوبية التي تستفيد من معالجة المدونات الموسمة بمختلف أنواعها:

- تطبيقات التحليل اللُّغوي الآلي (التحليل الصرفي، والنحوي، والدلالي).
- تطبيقات تحليل المشاعر الآلي (تحليل المشاعر والآراء التي تتضمَّنها البيانات والرموز اللُّغوية).
- الاستخلاص الآلي للمصطلحات وأسماء الأعلام: استخراج معلومات مُحدَّدة ومركبة من النصوص الخام غير المركبة، حيث يمكن التعامل مع هذه

المعلومـات بسـهولة آليًـا. وكذلـك اسـتخلاص أسـماء مُحـدَّدة مذكـورة في النـص، مثـل: أسـماء الأشـخاص والشـركات والمنظمـات وغيرهـا. وتُسـمَّى هـذه العمليـة باسـتخلاص أسـماء الأعـلام أو الكيانـات المُسـمَّاة.

- تطبيقـات التحليـل الـدلالي: حيـث تسـاعد الوسـوم الدلاليـة علـى تحديـد الأدوار والعلاقـات الدلاليـة داخـل البيانـات، وهـو أمـر جوهـري لتحسـين فهـم الحاسـوب للمعـاني الـتي تحملهـا البيانـات.
- تطبيقـات الترجمـة الآليـة: تُسـهم المدونـات اللُّغويـة في بنـاء نمـاذج حاسـوبية تُحسِّـن مـن مخرجـات الترجمـة الآليـة بـين اللُّغـات المختلفـة.
- تطبيقـات أخـرى متعـددة، مثـل: التقويـم الآلي للمقـالات، وتصنيفهـا، وتلخيصهـا، والتعـرُّف علـى الكاتـب.

ومـن المهـم الإشـارة إلى أن عمليـة توسـيم المدونـات مسـتقلة عـن بنـاء المدونـة اللُّغويـة الخـام، ففـي بعـض التطبيقـات اللُّغويـة، مـن المناسـب اسـتخدام المدونـة الخـام كمـا هـي دون توسـيم.

## مستويات التوسيم اللُّغوي للمدونات اللُّغوية النصية:

تتضمَّـن المعلومـات اللُّغويـة والوسـوم الـتي تُضـاف لنصـوص المدونـة مسـتويات لغويـة متعـددة، تشـمل كل مراحـل التحليـل اللُّغـوي بدايـةً مـن المسـتوى الصـوتي وانتهـاءً بمسـتوى تحليـل الاسـتعمال اللُّغـوي، وهـي كمـا يلـي:

- التوسـيم الصـوتي: يُقصَـد بـه إضافـة الرمـوز الـتي تشـير إلى العلامـات الصوتيـة في النـص، مثـل: علامـات الوقـف، أو النـبر والتنغيـم، وغيرهـا.

- التوسـيم الصـرفي: يتضمَّـن تقسـيم النـص وفقًـا للوحـدات الصرفيـة، وإضافـة الوسـوم الخاصـة بأقسـام الـكلام، أو التشـكيل الصـوتي الصـرفي للكلمـات.

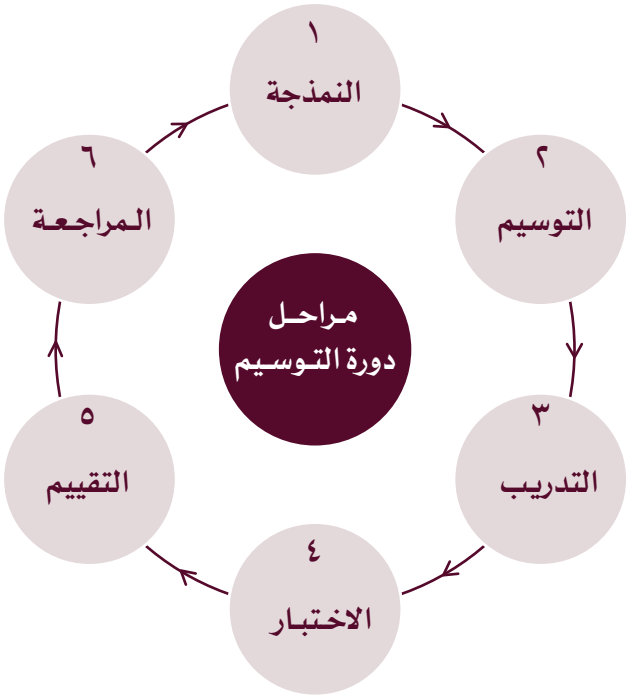
- التوسـيم النحـوي: يتضمَّـن إضافـة وسـوم العلاقـات النحويـة بـين مكونـات التراكيـب اللُّغويـة، مثـل: الحالـة الإعرابيـة للكلمـة في الجملـة، ونـوع علاقتهـا النحويـة بسـياقها.

- **التوسـيم الـدلالي:** يتضمَّـن التحليـل الـدلالي للنصـوص مـن خـلال ربـط الكلمـات بمفاهيـم وحقـول دلاليـة، وإضافـة معلومـات حـول العلاقـات الدلاليـة بـين الكلمـات، مثـل: الـترادف، والتضـاد، والاشـتراك اللفظـي، وغـير ذلـك.

- **التوسـيم التـداولي:** يشـمل إضافـة رمـوز خاصـة بتوضيـح الاسـتعمال اللُّغـوي في سـياقات متعـددة، إضافـةً إلى بيـان المقاصـد المحتملـة في الاسـتعمال اللُّغـوي.

## دورة التوسيم وخطواته:

مـن المهـم الإشـارة إلى أن عمليـة التوسـيم اليـدوي عمليـة مُكلفـة مـن ناحيـة الوقـت والجهـد والتكلفـة الماديـة؛ لمـا يترتـب عليهـا مـن توظيـف أشـخاص للقيـام بعمليـة التوسـيم. وقـد لا تتوفـر هـذه الإمكانيـات لـدى الكثيـر مـن الباحثـين. وعمليـة التوسـيم عبـارة عـن دورة تُعـاد بحسـب الحاجـة، وتمـرُّ عـبر عـدة مراحـل تتكـرر لتصحيـح الأخطـاء وتحسـين جـودة التوسـيم، وأحيانًـا لتغيـير الوسـوم المُسـتخدَمة. وحسـب منصـة عمـل MATTER (Stubbs & Pustejovsky, 2012)، تبـدأ عمليـة التوسـيم بنمذجـة لظاهـرة لغويـة بعـد رصدهـا، وبعدهـا يكـون التوسـيم وتدريـب نمـوذج آلي لذلـك، ثـم اختبـاره، وتقويـم النتـائج، ومراجعتهـا، وبنـاءً علـى ذلـك تُحـدَّث النمذجـة، وتسـتمر هـذه الـدورة حـتى يتـم التوصُّـل إلى النتيجـة المنشـودة. ويُوضِّـح الشـكل أدنـاه تسلسـل هـذه الـدورة.




الشكل (١): مراحل دورة التوسيم

**١- النمذجة:**

في هذه المرحلة يُوضَع وَصْف نظري لظاهرة لغوية. وتختلف النمذجة بشكل كبير بحسب الظاهرة اللُّغوية. فمثلًا: في حالة تصنيف رسائل البريد الإلكتروني إلى "مزعجة" و"غير مزعجة"، تكون النمذجة بوصف الوسمين "مُزعِج" و"غير مُزعِج"، بحيث يُحدَّد الوصف الذي ينطبق على كلٍّ من الفئتين، فمثلًا: يمكن وصف البريد المُزعِج بأنه "رسائل لا نرغب في قراءتها ونريد التخلُّص منها آليًّا"، والبريد غير المُزعِج بأنه "رسائل قد نرغب في قراءتها". **ويجري في هذه المرحلة:**

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

- البحث عن نظريات تصف الظاهرة اللُّغوية، مثلًا: في وسم أقسام الكلام، من المهم دراسة النظريات اللُّغوية حول ذلك.
- نمذجة الظاهرة اللُّغوية؛ وذلك بوَضْع الوصف العام لها، مثلًا: تحتوي بعض النصوص على أسماء أعلام (أو كيانات) تتميز بأنها تشير إلى كيانات حقيقة مُحدَّدة مثل الأشخاص والمنظمات وغيرها، ومن المفيد في كثير من التطبيقات تمييزها عن غيرها من الكلمات في النص.
- وَضْع مُخطط التوسيم بناءً على نمذجة الظاهرة اللُّغوية، بحيث يحتوي على الوسوم المستهدفة ووصف عام لها، وذلك بعد دراسة مخططات التوسيم السابقة المشابهة؛ بهدف الاستفادة منها أو تطويرها.
- البحث عن مدونات خام مناسبة، أو إنشاء مدونة جديدة في حالة عدم وجود مدونة مناسبة. وتُحدَّد مواصفات المدونة الخام بحيث تكون مناسبةً لغرض التوسيم، فمثلًا: النصوص الإخبارية الرسمية قد لا تكون مناسبةً لتطبيق تحليل المشاعر للمستهلكين. أيضًا، يجب مراعاة اختلاف النوع الأدبي واللهجات، وغيرها. ومن المهم أن تكون المدونة الخام شاملةً، وتعكس الواقع بحيث يمكن الاستفادة من التوسيم المضاف إليها لاحقًا.

**٢- التوسيم:**

في هذه المرحلة، تبدأ عملية التوسيم الفعلية، وذلك بإضافة الوسوم إلى المدونة الخام، وغالبًا ما يُشارك في هذه المرحلة عددٌ من الموسِّمين (المسؤولين عن تنفيذ التوسيم). وفي بعض التطبيقات، قد يكون من المناسب لرفع جودة التوسيم توسيم الأمثلة نفسها من أكثر من شخص. ويتفاوت عدد الموسِّمين بحسب الإمكانيات المتاحة سواء من حيث الموارد المادية أو توفُّر المُوسِّمين المُؤهَّلين. **وعادةً ما يجري في هذه المرحلة:**

- إعـداد دليـل التوسـيم: في المراحـل الأولى، يُوضَـع دليـل التوسـيم مـن الخـبراء بعـد توسـيم عينـة صغـيرة ووَضْـع مخطـط التوسـيم، وتسـتمر عمليـة تطويـر دليـل التوسـيم وتحسـينه مـع الوقـت.
- تحديـد حجـم المدونـة المُوسَّـمة: في هـذه المرحلـة تُحـدَّد الكميـة المُسـتهدَفة مـن التوسـيم، وتُقـاس غالبًـا بعـدد الكلمـات أو الجمـل أو القطـع النصيـة، وذلك بحسـب التطبيـق. والكميـة المسـتهدفة تخضـع لمعايـير عديـدة، مثـل: الميزانيـة الماليـة المُخصَّصـة للتوسـيم، وكذلـك طبيعـة المهمـة والتطبيـق، والوقـت المُسـتهدَف للانتهـاء مـن التوسـيم. وهـذه الأمـور تُحـدَّد مـن الخـبراء في المجـال، وبالرجـوع إلى مشـاريع توسـيم مشـابهة. ويُحـدَّث حجـم المدونـة بحسـب الحاجـة.
- اختيـار نوعيـة المُوسِّـمين وعددهـم: بعـد تحديـد منهجيـة التوسـيم، وإجـراء توسـيم مبـدئي مـن قِبـل الخـبراء، هنـا تُحـدَّد مؤهـلات المُوسِّـمين ونوعيتهـم بحسـب طبيعـة التطبيـق. فمثـلًا: في حالـة العمـل على مدونـات بلهجـة معينـة، مـن المهـم أن يكـون المُوسِّـم متحدِّثًـا بهـذه اللهجـة لهجـةً أُمًّـا. كذلـك في الحـالات الـتي تتطلـب توسـيمًا لغويًّـا متخصِّصًـا، كتحليـل البنيـة الصرفيـة أو المسـتويات المتقدمـة مـن المعالجـة اللغويـة، يجـب أن يتمتـع الموسِّـمون بخلفيـة لغويـة قويـة وخـبرة مناسـبة لطبيعـة المهمـة. أيضًـا في هـذه المرحلـة يُحـدَّد عـدد الموسـمين. ومـع تقـدم عمليـة التوسـيم، قـد يكـون مـن المناسـب تغيـير عـدد الموسـمين أو نوعيتهـم بحسـب نتيجـة العمـل. مثـلًا: قـد يـرى الخـبراء مـع الوقـت أن عـدد المُوسِّـمين لـكل مثـال أكـثر مـن الحاجـة، أو أن طبيعـة البيانـات تتطلـب موسِّـمين مـن خلفيـات معينـة؛ كأن ينتمـون إلى فئـة عمريـة أو يتحدثـون لهجـة معينـة، أو غـير ذلـك.
- تحديـد الأدوات المُسـتخدَمة للتوسـيم: بعـض الأدوات تُسـهِّل وتُسـرِّع عمليـة التوسـيم، وتُسـهِم في تقليـل الأخطـاء، بعضهـا بمقابـل مـادي والآخـر مجـاني. واختيـار

الأدوات المُستخدَمة يكون بحسب الميزانية المتوفرة، وحسب الإمكانيات التقنية، ومدى توافق الأداة مع متطلبات المشروع، وغيرها. أيضًا، من المهم التأكُّد من رُخص استخدام الأدوات المُستخدَمة، وهل تسمح باستخدامها لأغراض التوسيم؟ فبعض الأدوات مجانية إذا كانت لا تُستخدَم لأغراض تجارية، وقد لا تكون مناسبة لمَن يرغب في توسيم بيانات لأغراض تجارية.

- تخزين الأمثلة من المدونة الخام في صيغ مناسبة ومعالجتها: بعد تحديد الأدوات المُستخدَمة للتوسيم، تُحوَّل المدونة في شكلها الخام إلى الصيغة المناسبة التي يمكن التعامل معها في هذه الأدوات، مثلًا: لو ارتأى الخبراء استخدام جداول البيانات مثل Excel أو Google Sheets لتوسيم المدونة، فتُحوَّل المدونة في هذه الخطوة من شكلها الخام (مثل ملفات نصية) إلى جداول بيانات تسهل التوسيم. وفي هذه المرحلة أيضًا، لا يقتصر العمل على التحويل فقط، بل تُجرى أيضًا المعالجات اللازمة للبيانات، وفقًا لما تتطلبه المهمة من تهيئة وتنظيف وتنسيق، بما يضمن جاهزيتها الكاملة للمرحلة التالية من التوسيم.
- توزيع البيانات على الموسِّمين: وهنا تُوزَّع البيانات على الموسمين على دفعات، ومن المناسب -أيضًا- في كثير من التطبيقات توزيع العينة نفسها على أكثر من مُوسِّم بحيث تُوسَّم أكثر من مرة؛ لغرض رفع الجودة ومقارنة الأداء.
- جَمْع الوسوم من المُوسِّمين: بعد انتهاء كل دفعة، تُجمَع الوسوم من الموسِّمين.
- إجراء اختبارات الجودة: بعد جَمْع الوسوم تُجرَى اختبارات الجودة (سنُفصِّلها في فصل جودة التوسيم)، وتُكتشَف الأخطاء، ويُدرَّب الموسِّمون على تفاديها.
- تحديث دليل التوسيم بحسب الحاجة.
- متابعة عمل الموسِّمين من حيث عدد ساعات العمل، وكذلك مراقبة المصاريف من الميزانية المرصودة.

**٣- التدريب:**

في هذه المرحلة تُختار خوارزمية للتعلُّم الآلي، وتُدرَّب على البيانات المُوسَّمة.

**٤- الاختبار:**

بعد تدريب خوارزمية التعلم الآلي، يُختبَر النموذج المُدرَّب على بيانات الاختبار التي لم يتدرب عليها النموذج؛ وذلك بهدف تقييم أدائه وقياس قدرته على التعميم على بيانات جديدة.

**٥- التقييم:**

يُفرَّق في سياق تعلُّم الآلة بين تقييم النموذج وتقويم النموذج، رغم التقارب اللغوي بين المصطلحين. فتقييم النموذج (Model Evaluation) يشير إلى عملية قياس أداء النموذج المدرب باستخدام بيانات اختبار لم تُستخدم أثناء التدريب، وذلك من خلال مؤشرات محددة مثل الدقة (Accuracy)، والاستدعاء (Recall)، والدقة الإيجابية (Precision)، وF1-score، وغيرها، بحسب طبيعة المهمة. ويهدف هذا التقييم إلى التحقق من قدرة النموذج على التعميم والتعامل مع بيانات جديدة. أما تقويم النموذج (Model Rectification أو Calibration)، فيأتي كخطوة لاحقة تُتَّخذ عند الحاجة لتحسين أداء النموذج بناءً على نتائج التقييم. ويتضمن ذلك تصحيح الأخطاء أو التعديلات الممكنة مثل إعادة ضبط المعلمات (Hyperparameters)، أو تحسين جودة البيانات، أو تطبيق تقنيات معايرة، أو حتى تعديل بنية النموذج نفسه. باختصار، يُعنى التقييم بقياس الأداء، في حين يهتم التقويم بتحسينه ومعالجة جوانب القصور فيه.

**٦- المراجعة:**

إذا لم تكن نتائج التقويم مُرضيةً، تُعاد مراجعة العملية برمجتها بدءًا من تحليل الأخطاء التي وقع فيها النموذج. ففي بعض الأحيان، قد تعود هذه الأخطاء إلى قصور في خوارزمية التعلُّم الآلي نفسها، أو إلى نقص في حجم بيانات التوسيم، وقد تكون ناتجة

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

عـن مشـكلات في التوسـيم ذاتـه. في هـذه الحالـة، يصبـح مـن الضـروري مراجعـة البيانـات بعنايـة، وتعديلهـا، وتحسـين آليـة التوسـيم بمـا يضمـن رفـع جـودة الأداء. وتُكـرَّر هـذه الـدورة مـن التحليـل والتحسـين حـتى الوصـول إلى نتـائج تحقـق مسـتوى الرضـا المنشـود.

## التوسيم المباشر وغير المباشر:

الهـدف مـن عمليـة التوسـيم هـو إضافـة معلومـات لغويـة جديـدة إلى المدونـة الخـام، وهـذه المعلومـات قـد تكـون مباشـرة بحيـث تعطـي قطعـة مـن النـص -كجملـة أو كلمـة معينـة- وسـمًا محـددًا. أو أن تكـون غـير مباشـرة بحيـث تعطـي نصًّـا إضافيًّـا منفصـلًا عـن المدونـة الخـام.

- **التوسـيم غـير المباشـر:** والمقصـود بـه أن تضـاف الوسـوم بشـكل مسـتقل عـن النص، دون ارتبـاط مباشـر بعنصـر لغـوي مُحـدَّد داخلـه. ومـن الأمثلـة علـى ذلـك: الإجابـة عـن الأسـئلة؛ ففرضًـا لـو كانـت المدونـة الخـام تتألـف مـن مقـالات موسـوعية، والوسـوم المطلوبـة هـي إجابـة عـن أسـئلة يمكـن اسـتخلاصها مـن محتـوى النـص؛ ففـي هـذه الحالـة يعـبر التوسـيم تعبـيرًا غـير مباشـر؛ لكونـه لا يشـير إلى قطعـة معينـة في النـص، بـل قـد يتطلـب الأمـر قـراءة أجـزاء متفرقـة مـن النـص. أيضًـا مـن الأمثلـة الأخـرى علـى التوسـيم غـير المباشـر: تلخيـص النصـوص؛ ففـي هـذه الحالـة لا يكـون الوسـم مرتبطًـا بجـزء معـين مـن النـص بحيـث يُوسَّـم كل جـزء علـى حـدة؛ بـل يتطلـب كتابـة الملخـص بشـكل مسـتقل، معتمـدًا علـى الفهـم العـام للمحتـوى، دون الإشـارة الصريحـة إلى مواقـع محـددة في المدونـة الخـام.
- **التوسـيم المباشـر:** في التوسـيم المباشـر تضـاف وسـوم تحـوي معلومـات إضافيـة لقطـع مـن النـص مباشـرة، بحيـث يُعـدّل عليـه، أو يكـون متوازيًـا معـه. ويُقصـد بذلـك أن تكـون المعلومـات المضافـة مرتبطـة ارتباطًـا مباشـرًا وواضحًـا بعنصـر لغـوي معـين

داخل المدونة. مثلًا: في حالة توسيم الكلمات بأقسام الكلام الخاصة بكل كلمة، هنا نحتاج إلى أن نضيف لكل كلمة على حدة الوسم الخاص بها؛ ولذلك يكون التوسيم مباشرة على مستوى النص وليس بشكل منفصل عنه.

## مُسوِّغات بناء المعايير:

مع التقدُّم الكبير في نماذج الذكاء الاصطناعي التوليدي، ومعالجة اللُّغة الطبيعية بشكل عام، زادت الحاجة إلى وجود بيانات موسَّمة لأغراض عدة، منها تقييم هذه النماذج، ويتطلب ذلك أن تكون هذه البيانات مبنيةً بطريقة معيارية يمكن الاعتماد عليها. وتأتي هذه الوثيقة لتُقدِّم إطارًا شاملًا للمعايير الخاصة بتوسيم المدونات اللُّغوية العربية؛ وذلك لأسباب عدة يمكن إجمالها على النحو الآتي:

**١- الحاجة إلى تحسين جودة البيانات اللُّغوية الموسّمة:**

إن توسيم المدونات اللُّغوية هو عمليةٌ دقيقةٌ تنطوي على إضافة مستويات مختلفة من المعلومات والوسوم إلى النص أو الصوت؛ مما يجعله مؤهلًا لتطبيقات متقدمة في مجالات عدة، مثل: التحليل النحوي والدلالي، واستخلاص المعلومات، والترجمة الآلية. وبغياب معايير واضحة تُنظِّم عملية التوسيم، قد تتفاوت جودة المخرجات والوسوم بشكل كبير، وهو ما يؤثر سلبًا في دقة النماذج المبنية على تلك البيانات. لذا، فإن وضع معايير منهجية أمر في غاية الأهمية؛ وذلك لضبط جودة البيانات، وتعزيز مصداقية نتائج الأبحاث والتطبيقات الحاسوبية.

**٢- تنوُّع المنهجيات وضعف توحيد الجهود:**

في المشاريع اللُّغوية العربية، غالبًا ما يُلاحظ تنوعٌ في الأساليب والأدوات المُستخدَمة لتوسيم المدونات، نتيجةً لعدم وجود إطار مُوحَّد يتبنَّاه الباحثون والمُطوِّرون. هذا

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

التشتُّت المنهجي يعيق تكامل الجهود وتبادل النتائج، ويؤدي إلى صعوبة إعادة استخدام البيانات الموسّمة أو دمجها في مشاريع جديدة. لذا، تسعى هذه الوثيقة إلى توحيد الرؤى والمعايير، بما يسمح بتقليل إهدار الوقت والموارد، ويُعزّز من قابلية الاستفادة من المدونات الموسّمة في أعمال بحثية متعددة.

**٣- تحقيق الاستدامة والقابلية للتحديث:**

تُعَدُّ المدونات الموسّمة مادةً حيَّةً تتطوَّر بتطوُّر اللُّغة والاحتياجات البحثية والتقنية. وبغياب معايير واضحة تُحدِّد كيفية توسيم النصوص وصيانتها وتحديثها؛ تظلُّ عملية التوسيم عُرضةً للارتجال واجتهادات شخصية قد لا تواكب التغيّرات اللُّغوية أو التطورات التكنولوجية المتسارعة. لذا، فإن اعتماد معايير ثابتة وقابلة للتطوير يُمكِّن من تحديث المدونات الموسّمة دوريًّا، مع الحفاظ على اتساق وسومها وانتظام بنيتها الداخلية.

**٤- ضمان الاتساق في مستويات التوسيم اللُّغوي:**

يشمل التوسيم مستويات لغوية عدة (الصرفي، والنحوي، والدلالي، والتداولي، إلخ)، ولكل مستوى معايير ووسوم مُتخصِّصة. وغياب إطار شامل يُحدِّد الأسس المُوحَّدة لتعيين الوسوم وتنظيمها قد يؤدي إلى التباس في فهم البيانات أو دمجها، إذ قد يختلف مفهوم الوسم الواحد من مشروع لآخر. من هنا تنبع الحاجة إلى وثيقة مرجعية موحدة توفّر توصيفًا دقيقًا للوسوم، وتُسهِم في توحيد الرموز والمفاهيم؛ بما يضمن اتساق التوسيم وقابلية تبادل البيانات بين الباحثين والمطورين عبر مختلف المبادرات والمشروعات.

**٥- تعزيز موثوقية التطبيقات الحاسوبية:**

ترتبط دقّة وموثوقية العديد من تطبيقات المعالجة الآلية للغة بمدى جودة التوسيم؛ فعلى سبيل المثال، يعتمد تصحيح النص الآلي، واستخراج الكيانات المُسمَّاة،

والترجمة التفاعلية، على مدى دقة تحديد أقسام الكلام والعلاقات النحوية والدلالية. وبإقرار معايير مُحكَمة، نضمن أن تكون الوسوم متناسقةً ومتوافقةً مع الاشتراطات الفنية للنماذج الحاسوبية؛ وهو ما يجعل هذه التطبيقات أكثر اعتماديةً.

**٦- مواجهة تحديات اللُّغة العربية في مجال اللسانيات الحاسوبية:**

تمتاز اللُّغة العربية بسمات فريدة (كالتصريف، والإعراب، والتنوع اللهجي) وغيرها من السمات التي تجعل التعامل معها أكثر تعقيدًا مقارنةً بلغات أخرى. وهذا يُبرز الحاجة الملحّة إلى معايير دقيقة تُراعي هذا التعقيد وتتناغم مع طبيعة العربية وبنيتها الغنية. فوجود مثل هذه المعايير لا يسهم فقط في سدِّ الفجوات المنهجية التي يواجهها الباحثون عند العمل على النصوص العربية الموسّمة، بل يُمثّل أيضًا خطوةً محوريةً نحو تطوير تطبيقات الذكاء الاصطناعي باللغة العربية، ورفع مستوى دقتها وكفاءتها بما يتناسب مع خصوصية هذه اللغة.

## المبادئ التي قامت عليها هذه المعايير:

- الشمول والتدرُّج: تغطية مختلف مراحل التوسيم اللُّغوي؛ ابتداءً من تحديد الهدف من التوسيم، إلى اختيار المدونة المناسبة، مرورًا بالمراحل المختلفة، مثل: إعداد دليل التوسيم، وتدريب المُوسِّمين.
- القابلية للتطبيق والتطوير: صياغة معايير مرنة يمكن تطبيقها في بيئات متنوعة، وقابلة للتوسُّع أو التكيُّف مع التقنيات الجديدة.
- التوافق مع الممارسات العالمية: السعي إلى مواءمة المعايير المطروحة مع الوثائق والمبادرات الدولية في مجال الترميز والتوسيم، بحيث تُسهِّل تبادل المدونات واستيراد الخبرات من المجالات البحثية المختلفة.
- الوضوح والتوثيق: ضمان أنَّ عملية التوسيم موثقة في كل مراحلها، مع توفير

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

أدلـة إرشـادية ووسـوم مُتَّفـق عليهـا، بحيـث تُقلِّـل مـن احتماليـة سـوء الفَهْـم أو التضـارب في اسـتخدام هـذه الوسـوم.

- المراجعـة والتقييـم المسـتمر: إتاحـة الفرصـة لتقييـم المعايـير بشـكل دوري، واسـتقبال التغذيـة الراجعـة مـن المجتمع البحثي؛ بهـدف تطويرهـا وتحسـينها على ضـوء المخرجـات والتجـارب العمليـة.

تأسيسًـا على مـا سـبق، يتضـح أنَّ وجـود معايـيرمُحـدَّدة لتوسـيم المدونـات العربيـة يُعَـدُّ ضـرورةً يفرضهـا واقـع البحـث والتطويـر في مجـال اللسـانيات الحاسـوبية وتقنيـات الـذكاء الاصطناعـي. ويُعـوَّل علـى هـذا الإطـار المعيـاري في توحيـد الجهـود، وتعزيـز قابليـة تبـادل المدونـات الموسّـمة بـين المؤسسـات البحثيـة؛ ممـا يفتـح البـاب أمـام مشـاريع مسـتقبلية تتسـم بالفاعليـة والاسـتدامة في دراسـة اللُّغـة العربيـة وتطبيقاتهـا التقنيـة.

## الدراسات والممارسات السابقة:

شـهد مجـال توسـيم المدونـات اللُّغويـة تطـوُّرات تاريخيـة مهمـة، حـتى أصبـح يُمثِّـل حجـرَ الزاويـة في مجـالي معالجـة اللُّغـة الطبيعيـة واللسـانيات الحاسـوبية؛ وذلـك لأهميـة المدونـات الموسّـمة في تطويـر نمـاذج الـذكاء الاصطناعـي الـتي تعتمـد اعتمـادًا رئيسًـا علـى هـذا النـوع مـن البيانـات.

يسـتعرض Pustejovsky & Stubbs (2012) وكذلـك Ide (1998) تاريـخ لسـانيات المدونـة، حيـث بـدأ العمـل في المدونـات اللُّغويـة في الخمسـينيات مـن القـرن المـاضي، وكان التركيـزحينها علـى بنـاء مدونـات لغويـة خـام؛ وذلـك بجَمْـع المـواد اللُّغويـة بشـكلها الطبيعـي مـن دون إضافـات لغويـة جوهريـة. وقـد شـهد عقـد التسـعينيات انطلاقـة واسـعة في ميـدان توسـيم المدونـات، حيـث انصبَّـت الجهـود آنـذاك علـى التحليـل النحـوي والتصنيـف الدقيـق لأقسـام الـكلام. ومـع مـرور الوقـت، اتسـع نطـاق التوسـيم ليشـمل مسـتويات لغويـة

متعـددة، بـل وتجـاوز ذلـك ليغطـي أيضًـا الجوانـب شـبه اللغويـة، مثـل النصـوص المصاحبـة للصـوت، والبيانـات السـمعية والبصريـة، بمـا في ذلـك محتـوى الفيديـو.

ركَّـزت الأعمـال المبكـرة في توسـيم المدونـات اللُّغويـة علـى اللُّغـة الإنجليزيـة، وخاصـة مـا يخـصُّ الظواهـر اللُّغويـة النحويـة بهـدف فهـم التراكيـب اللُّغويـة؛ حيـث قـام الخـبراء اللُّغويـون بتوسـيم النصـوص الإنجليزيـة وفقًـا لنظريـات وأطـر نحويـة محـددة؛ ممـا أرسى الأسـاس لتحليـل نحـوي منهـجي يمكـن البنـاء عليـه في التطبيقـات الحاسـوبية. ولعـل أبـرز مشـروعات توسـيم المدونـات المبكـرة عالميًّـا هـو مشـروع "بنـك پـين الشـجري" Penn Treebank، الـذي نُفِّـذَ بجامعـة بنسـلفانيا بالولايـات المتحـدة (Marcus et al., 1993)، وهـو مدونـة موسـومة بأقسـام الـكلام والتراكيـب النحويـة تضـمُّ نحـو 4.5 ملايـين كلمـة باللُّغـة الإنجليزيـة. وقـد شـكّلت هـذه المدوّنـة الأسـاس المنهـجي الـذي انطلقـت منـه الجهـود اللاحقـة في إرسـاء معايـير توسـيم المدونـات اللغويـة، سـواء مـن حيـث التأصيـل النظـري أو مـن حيـث التطبيقـات العمليـة.

ومع انتشـار الشـبكة العالميـة العنكبوتيـة في بدايـة الألفيـة الثالثـة، أصبحـت البيانـات متوفـرة بشـكل أكـبر لاسـتخدامها في النمـاذج الإحصائيـة لعـدة تطبيقـات، وفي مقدّمتهـا الترجمـة الآليـة. وقـد بـرزت في تلـك الفـترة العديـد مـن المدونـات الموسّـمة في مختلـف أنحـاء العالـم. ومـع انتقـال كثـير مـن النمـاذج الحاسـوبية في المعالجـة الآليـة للغـة مـن المنهـج القائـم علـى القواعـد اللُّغويـة المحوسـبة إلى منهـج التعلـم الآلي المبـني علـى نمـاذج وتقنيـات الـذكاء الاصطناعـي؛ ازدادت الحاجـة إلى مدونـات لغويـة موسـومة لتدريـب نمـاذج الـذكاء الاصطناعـي عليهـا وتحسـين كفاءتهـا التقييميـة. ونتيجـة لذلـك، كـثرت الدراسـات والمحـاولات التطبيقيـة الخاصـة بتوسـيم المدونـات، وضبـط عمليـة اختيـار الوسـوم وتحليلها، وتطويـر وثائـق مرجعيـة للتوسـيم في لغـات ومجـالات متعـددة.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

ويُشار إلى أن المعايير الخاصة بتوسيم المدونات يمكن تقسيمها لنوعين حسب الغرض الذي صُمِّمت بدايةً من أجله: الأول، وهو الأكثر شيوعًا، معايير تُعَدُّ "بحكم الأمر الواقع" (de facto standards)، إذ لم تُصمَّم في الأصل لتكون مرجعًا عامًّا؛ وإنما اكتسبت صفة المعيار نتيجةً لانتشار المدونة التي طُبّقت فيها، وسبقها الزمني في المجال، مما أكسبها مكانةً مرجعيةً ضمن المجتمع العلمي، رغم أنها لم تهدف في الأساس إلى توجيه مشروعات توسيم أخرى. أما النوع الثاني، فهو المعايير التي وُضعت خصيصًا لتكون مرجعًا موحّدًا يُعزز التوافق بين مشروعات التوسيم المختلفة. وفي هذا السياق، تُعدّ المعايير المعتمدة في مشروع "بنك بن الشجري" (Penn Treebank) مثالًا على النوع الأول؛ إذ اكتسبت صفة المعيار بفضل ريادتها وانتشارها الواسع، وليس لأنها صُمّمت لغرض التوافق بين المشاريع.

وفيما يخص النوع الثاني من المعايير، لعل أبرز الجهود العالمية لوضع معايير للمدونات وتوسيمها هو مشروع معيار ترميز المدونات Corpus Encoding Standard CES (Ide, 1998) الذي أُطلِق تحت مبادرة ترميز النصوص The Text Encoding Initiative TEI، وهو تطبيق للمعايير التي طُرحت في مبادرة ترميز النصوص TEI على المدونات اللُّغوية. ويُشار إلى أن معيار ترميز المدونات CES يشمل المدونات الخام والمدونات الموسّمة، مع وجود فصلين مستقلين لكلٍّ منهما في هذا المعيار. كذلك، تُمثِّل دراسة (Garside & Leech, 1997) نموذجًا للمحاولات المبكرة لتأصيل وتنظيم عمليات التوسيم للمدونات اللُّغوية. ومن الجدير بالتنويه أن المعايير من النوع الثاني قد تأثرت تأثرًا بالغًا بوجود النوع الأول، حيث أسهمت الوثائق والمستندات المنشورة للمدونات الموسّمة مثل بنك بين الشجري Penn Treebank في تطوير معايير توسيم المدونات؛ وذلك لأن هذه الوثائق تحتوي عادةً على معلومات قيمة حول عملية

التوسـيم، مثـل: اختيـار الوسـوم المناسـبة، وطريقـة كتابـة المسـتندات، وغيرهـا. وعلـى ذلـك، فـإن المعايـير مـن النـوع الثـاني (وهـو مـا تُقدِّمـه هـذه الوثيقـة) تعتمـد بشـكل رئيـس علـى مـا تراكـم مـن خـبرات ومعايـير فعليـة في النـوع الأول.

ولاحقًـا، أدركـت منظمـة المعايـير العالميـة International Organization for Standardization (الأيـزو ISO) أهميـة وضـع معايـير للمدونـات وتوسـيمها، ثـم طـوَّرت عـدة معايـير في هـذا المجـال؛ لعـل أهمهـا "إطـار عمـل التوسـيم اللُّغـوي (Linguistic Annotation Framework – Ide & Suderman, 2014) (LAF)، الـذي نُشـر رسـميًّا في عـام 2012. ويُعـدُّ هـذا الإطـار نموذجًـا عامًّـا لتصميـم مشـروعات التوسـيم، إذ يتميَّـز بمرونتـه العاليـة؛ ممـا يتيـح تطبيقـه علـى مختلـف المسـتويات اللغويـة.

كمـا نشـرت منظمـة الأيـزو عـددًا مـن المعايـير الأخـرى ذات الصلـة، بعضهـا يركّـز علـى مسـتويات محـددة مـن التوسـيم، مـن ذلـك معيـار "تفريـغ اللغـة المنطوقـة" (Transcription of Spoken Language)، الـذي صـدر في عـام 2016، ويُعـنى بوضـع أطـر منهجيـة لتفريـغ المدونـات الصوتيـة وتحويلهـا إلى نصـوص مكتوبـة وفـق ضوابـط موحّـدة، تضمـن الدقـة والتوافـق في معالجـة اللغـة المنطوقـة ضمـن مشـروعات المدونـات اللغويـة.

وأمـا فيمـا يتعلّـق بتوسـيم المدونـات اللغويـة العربيـة، فقـد انطلقـت الجهـود الأولى مُركّـزة علـى التحليـل الصـرفي والنحـوي، ولا سـيما في سـياق النمـاذج الحاسـوبية المُدرَّبـة علـى تصنيـف أقسـام الـكلام. ومـع مـرور الوقـت، توسَّـعت هـذه الجهـود تدريجيًّـا لتغطـي مسـتويات لغويـة أوسـع، وتشـمل أيضًـا معلومـات لغويـة دقيقـة، بـل وتتجاوزهـا أحيانًـا إلى مـا يُعـرف بالبيانـات مـا وراء اللغويـة.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

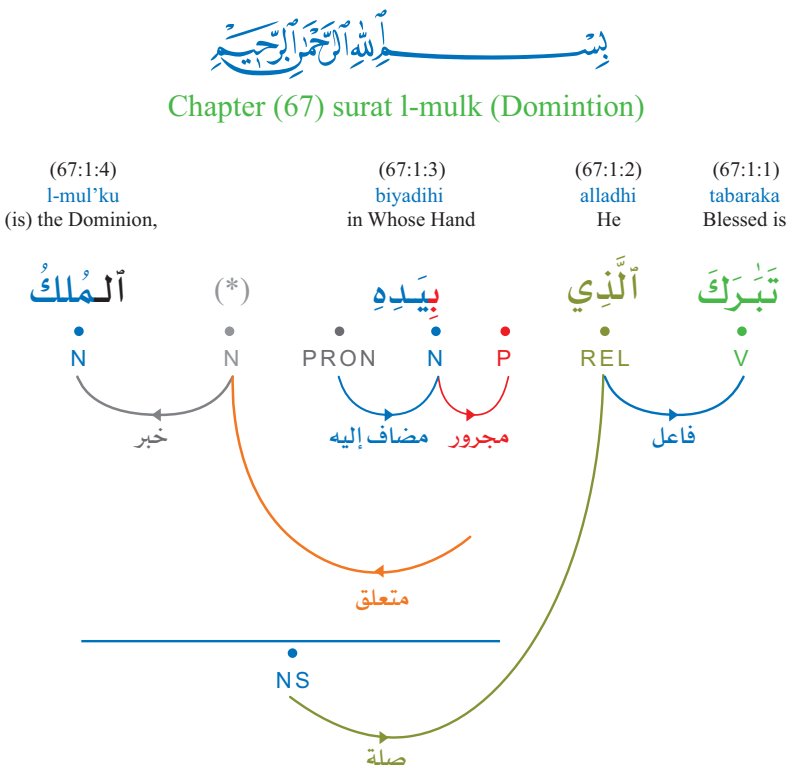
وعلـى الرغـم مـن توفـر عـدد مـن المدونـات اللُّغويـة العربيـة الموسّـمة، إلا أنـه يوجـد نقـصٌ في المعلومـات الخاصـة بمعاييـر تصميمهـا ومـدى اعتمادهـا علـى دليـل أو إطـار نظـري للتوسـيم. وعنـد اسـتعراض الأدبيـات السـابقة في توسـيم المدونـات العربيـة نلاحـظ عـددًا مـن المحـاولات المحـدودة، منهـا مؤلَّفـات قدَّمـت تعريفًـا بالمدونـات اللُّغويـة ومراحـل بنائهـا وتوسـيمها، علـى سـبيل المثـال مـا جـاء في كتـاب "المدونـات اللُّغويـة العربيـة: بناؤهـا وطرائـق الإفـادة منهـا" (العصيـمي، ٢٠١٥)، الـذي يضـمُّ سـتة مباحـث تناقـش القضايـا الرئيسـة في لسـانيات المدونـة، ويتميَّـز الكتـاب بكونـه أول الإصـدارات فيمـا يتعلـق بلسـانيات المدونـة، كمـا أنـه مـن أوائـل إصـدارات مركـز الملـك عبـد الله بـن عبـد العزيـز الـدولي لخدمـة اللُّغـة العربيـة، ويركِّـز بشـكل خـاص علـى المدونـات باللُّغـة العربيـة. حيـث يناقـش المبحـث الثالـث (تصميـم المدونـات اللُّغويـة وبناؤهـا)، وقـد ورد في الكتـاب بعـض المعاييـر ذات العلاقـة ببنـاء وتوسـيم المدونـات، مـع التركـيز علـى بنـاء المدونـات الخـام. ويُلاحـظ عـدم التفريـق بـين المعاييـر الخاصـة ببنـاء المدونـة الخـام والمعاييـر الخاصـة بتوسـيم المدونـة اللُّغويـة. كمـا صـدرت مُؤلَّفـات أخـرى تتنـاول التوسـيم اللغـوي للغـة العربيـة، مـن بينهـا "دليـل توسـيم البيانـات اللغويـة" (الثبيـتي، ٢٠٢١)، الـذي يسـتعرض مراحـل التوسـيم وطرقـه، وكذلـك ورقـة "وسـم المدونـات اللغويـة: المفهـوم والمجـالات" (الفيفـي، ٢٠٢٣)، الـتي تـبرز مفهـوم وسـم المدونـات وتُفـرّق بـين أنـواع الوسـوم الـتي يمكـن إضافتهـا إلى المدونـات الخـام.

وإضافـةً إلى الدراسـات المسـحية والوصفيـة للمدونـات اللُّغويـة وتوسـيمها باللُّغـة العربيـة، هنـاك عـدد مـن أدلـة ووثائـق التوسـيم اللُّغـوي للمدونـات العربيـة، مثـل مجموعـة وثائـق مختـبر كامـل CAMeL Lab Guidelines بجامعـة نيويـورك أبـو ظـبي[1]، الـتي تشـمل جوانـب لغويـة متعـددة في توسـيم المدونـات العربيـة، مثـل: التفريـغ الصـوتي والتحليـل الصـرفي لعـدد مـن اللهجـات العربيـة.

(١) https://camel-guidelines.readthedocs.io/en/latest/

كذلك، هناك العديد من المدونات اللُّغوية العربية الموسّمة، والبعض منها نشر أدلة التوسيم وتفاصيله، ونشير هنا باختصار لبعض الأمثلة للمدونات اللُّغوية الموسّمة، مع ملاحظة أن معظمها مقتصرة على مستوى التوسيم الصرفي والنحوي.

ومن ذلك: مدونة القرآن الكريم العربية The Quranic Arabic Corpus، وهي مدونة موسومة بمعلومات صرفية ونحوية ودلالية لجميع كلمات القرآن الكريم. وأعدَّها باحثون في جامعة ليدز Leeds في المملكة المتحدة[٢]. وتُقدِّم المدونة ترجمةً للكلمات والعبارات في القرآن الكريم من اللُّغة العربية إلى الإنجليزية. ويُوضِّح (الشكل ٢) مثالًا للبنك الشجري لبداية سورة الملك. (الشكل من الموقع الخاص بالمدونة على الإنترنت).




الشكل (٢): التوسيم الصرفي النحوي أو البنك الشجري للآية الأولى من سورة الملك في مدونة القرآن الكريم.

(٢) https://corpus.quran.com

ومنها كذلك البنك الشجري العربي Arabic Treebank (Maamouri et al., 2004): وهو مدونة موسومة من عمل جمعية البيانات اللُّغوية Linguistics Data Consortium LDC في جامعة بنسلفانيا في الولايات المتحدة. وتحتوي المدونة على وسوم لبيانات باللُّغة العربية؛ تشمل أقسام الكلام والتحليل الصرفي والنحوي، وجُمِعت نصوص المدونة من مصادر ومواقع إخبارية متفرقة.

وهناك بنوك شجرية أخرى للغة العربية، مثل: مدونة كاتب CATiB التي طوَّرتها مجموعة بحثية في جامعة كولومبيا في الولايات المتحدة. وكذلك مدونة كارانير-CAraN ER لأسماء الأعلام: وهي مدونة من تطوير مدينة الملك عبد العزيز للعلوم والتقنية، موسومة بأسماء الأعلام (شخص، منظمة، لقب... إلخ)، وتغطي الفترة الزمنية من بداية 2019 إلى نهاية 2021، وقد جُمعت بياناتها من عدة مواقع إخبارية في المملكة العربية السعودية، مع عناية خاصة برصد أسماء الأعلام الدارجة في السياق المحلي، مثل: أسماء المدن والمحافظات.

ومن بين المدونات المتخصِّصة في توسيم أسماء الأعلام، تبرز "مدونة وجود (Wojood)" (Jarrar et al., 2022) التي طوّرها فريق بحثي في جامعة بيرزيت بفلسطين. وتُعنَى هذه المدونة بتوسيم أسماء الأعلام في نصوص مكتوبة باللغة العربية الفصحى، إلى جانب بعض النصوص باللهجات العربية الدارجة. وتتميَّز المدونة بنظام توسيم دقيق يتيح التوسيم المتداخل، حيث يمكن إسناد أكثر من وسم لاسم علم واحد في حال احتوائه على أكثر من كيان معرفي. فعلى سبيل المثال، في اسم العلم "بنك الرياض"، يُوسَّم الاسم كاملًا باعتباره "شركة تجارية"، في حين تُوسَّم كلمة "الرياض" ضمنه أيضًا كـ "اسم مدينة"، مما يُبرز مرونة التوسيم وقدرته على تمثيل البنية المعرفية المتداخلة في النصوص العربية.

يتبـين ممـا سـبق أن الدراسـات والممارسـات السـابقة -سـواء المؤلَّفـات في وصـف المدونـات وطـرق بنائهـا وتوسـيمها، أو أدلـة التوسـيم المنشـورة للمدونـات اللُّغويـة العربيـة- لا تُوضِّـح بشـكل كافٍ المعايـير الـتي ينبغـي اسـتخدامها عنـد توسـيم المدونـات اللُّغويـة. وبنـاءً علـى ذلـك، تسـعى هـذه الوثيقـة إلى جمـع أهـم المعايـير العالميـة الـتي تخـصُّ توسـيم المدونـات، مـع التركيـز علـى المعايـير الخاصـة باللُّغـة العربيـة؛ وذلـك عـن طريـق مراجعـة مـا اسـتقرَّ عليـه العمـل في المدونـات العالميـة مـن جهـة، والمدونـات العربيـة مـن جهـة أخـرى.

## نطاق المعايير وحدودها:

تشـمل معايـير هـذه الوثيقـة جميـع مراحـل توسـيم البيانـات اللُّغويـة بدايـةً مـن الخطـوات الأولى الخاصـة بتحديـد هـدف التوسـيم، مـرورًا بمرحلـة تدريـب الموسِّـمين وتأهيلهـم وكتابـة دليـل التوسـيم اللُّغـوي؛ وانتهـاءً بتخزيـن المدونـة الموسّـمة، وتدويـن وتوثيـق عمليـة التوسـيم بـكل تفاصيلهـا، ومـن أهـم الأمـور الـتي ينبغـي مراعاتهـا في أثنـاء عمليـة التوسـيم اللُّغـوي مـا يـأتي:

- **وضـوح الهـدف:** بحيـث يكـون هـدف التوسـيم واضحًـا ودقيقًـا؛ وذلـك لتحديـد نـوع الوسـوم المُسـتهدَفة وطبيعتهـا، والطُّـرق المُثلَـى لإضافتهـا للمدونـة وفقًـا للتطبيـق اللُّغـوي الـذي تُعَـدُّ مـن أجلـه.
- **توظيـف الموسّـمين وتدريبهـم:** مـن المهـم تقديـم دورات وورش عمـل للمشـتركين في توسـيم بيانـات المدونـات اللُّغويـة؛ لضمـان جـودة العمـل البشـري في التوسـيم اللُّغـوي، وتجنُّـب عيـوب التوسـيم غـير الآلي وسـلبياته.
- **تطويـر وبنـاء وثائـق التوسـيم:** ينبغـي صياغـة أدلـة توسـيم واضحـة، تُحـدِّد المنهجيـة والمعايـير المتبعـة في عمليـة التوسـيم اللُّغـوي، وتكـون المرجـع المُعتمـد عليـه في أداء مهمـات التوسـيم مـن جميـع المشـاركين.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

- **تحسين أدلة التوسيم:** من الضروري أن تخضع أدلة التوسيم لعمليات تحديث وتطوير مستمرة؛ بهدف الاستجابة للتحديات المتجددة والملاحظات العملية التي تظهر أثناء الاستخدام؛ مما يسهم في رفع كفاءتها ودقتها وتعزيز مواءمتها مع المستجدات اللغوية والتقنية.
- **ترميز التوسيم:** يُراعى في عملية التوسيم اعتمادُ نُظُم ترميز قياسية تتيح قراءة الوسوم ومعالجتها بفاعلية، سواء آليًّا من قِبل البرمجيات، أو بشريًّا من قِبل الباحثين والمختصين؛ مما يضمن سهولة الاستخدام وتكامل البيانات في البيئات التقنية المختلفة.
- **تخزين البيانات وإدارتها:** وذلك لضمان التخزين الآمن وسهولة الوصول إلى المدونة الموسّمة لاحقًا.

## قائمة المعايير:

يمكن تصنيف معايير توسيم المدونات اللغوية العربية إلى سبعة مجالات رئيسة، تُغطّي المراحل الزمنية المختلفة لمشروع التوسيم، بدءًا من التخطيط وحتى النشر والتشغيل. وتشترك الستة الأولى منها مع المعايير العامة المتبعة في اللغات الأخرى، بينما يُعد المجال السابع مُخصَّصًا للغة العربية؛ نظرًا لما تتفرّد به من خصائص لغوية على المستويات المختلفة. وتشمل هذه المجالات ما يأتي:

١. **تحديد الهدف من التوسيم:** بدايةً ينبغي تحديد الهدف من بناء مدونة لغوية عربية موسومة (كأن تكون مُوجَّهة في تصميم نموذج لغوي قادر على التحليل الصرفي للغة أو لهجة معينة)، وتحديد الوسوم المستهدفة، وتحديد الظاهرة اللُّغوية المستهدفة (مثل: تحديد أقسام الكلام للغة أو لهجة محددة)، وكذلك الإطار النظري الذي يصف هذه الظاهرة اللُّغوية من منظور لساني أو نفسي أو

غـيره. ويُراعـى كذلـك تحديـد الميزانيـة المتوفـرة وتكاليـف التوسـيم المتوقعـة، وتحديـد المدونـة الخـام المناسـبة لذلـك ومواصفاتهـا، ومراعـاة أن تكـون متنوعـةً وشـاملةً؛ لضمـان تمثيـل واسـع لظواهـر اللغـة المسـتهدفة.

٢. **بنـاء دليـل ومنصـة التوسـيم:** دليـل التوسـيم هـو المرجـع الرئيـس للمسـتخدمين والموسِّـمين في التعامـل مـع المدونـة الموسّـمة؛ إذ يتـولى تحديـد الوسـوم المسـتهدفة، وتوضيـح آليـة التوسـيم، وتقديـم أمثلـة توضيحيـة ضمـن سـياقات لغويـة واضحـة. كمـا يبـيّن كيفيـة التعامـل مـع الحـالات الإشـكالية الـتي قـد تختلـف فيهـا آراء الموسِّـمين، ويضـع المعايـير المتفـق عليهـا لحسـمها. ومـن الضـروري أن يكـون هـذا الدليـل شـاملًا ودقيقًـا، يشـرح بالتفصيـل طريقـة اسـتخدام منصـة التوسـيم والأدوات المصاحبـة لهـا، بمـا يضمـن توحيـد أسـاليب العمـل وتيسـير التوسـيم علـى نحـو منهـجي ومنظّـم.

٣. **تدريـب الموسِّـمين واختيارهـم:** مـن أهـم المعايـير الـتي تسـهم في بنـاء مدونـة موسـومة بجـودة عاليـة هـي اختيـار الموسـمين المناسـبين، وكذلـك تدريبهـم وتحسـين أدائهـم بشـكل مسـتمر. يبـدأ اختيـار الموسـمين بتحديـد الخـبرات المطلوبـة وخلفياتهـم الثقافيـة واللُّغويـة وتنوُّعهـا، ثـم إخضـاع المُرشَّـحين لاختبـارات معياريـة؛ للتأكـد مـن قدرتهـم علـى تنفيـذ مهـام التوسـيم وفـق المعايـير المحـددة. ويلـي ذلـك تدريـب عملـي منهـجي يتضمَّـن مراجعـة دوريـة لأعمالهـم، وتقديـم تغذيـة راجعـة لتحسـين الأداء. كمـا يُستحسـن الحفـاظ علـى اسـتمرارية الفريـق المشـارك في التوسـيم لأطـول فـترة ممكنـة، مـع توفـير بيئـة عمـل مُحفِّـزة للحـدّ مـن معـدلات التبديـل؛ لمـا لذلـك مـن أثـر مباشـر في اتسـاق الوسـوم وثبـات جـودة المدونـة الموسّـمة.

٤. **جـودة التوسـيم:** تُطبَّـق المعايـير الـتي تسـهم في رفـع جـودة المدونـة الموسّـمة، مثـل أن تكـون متسـقة بـدون اختـلاف كبـير، وكذلـك تحديـد الأمثلـة التي يحصـل فيهـا اختلاف

بين الموسمين، ومراجعة دليل التوسيم بصورة دورية لتحديثه ومعالجة أوجه الغموض أو الالتباس؛ مما يسهم في تعزيز وضوح المعايير وتقليل التناقضات داخل المدونة الموسّمة.

٥. **تخزين المدونة الموسّمة:** بعد اكتمال توسيم المدونة، تأتي مرحلة تخزينها إلكترونيًّا حتى يسهل للمستخدمين الوصول إليها. وينبغي في هذه المرحلة تطبيق المعايير المتبعة لتخزين المدونة الموسّمة خاصة من ناحية الترميز ونوع الملفات. وكذلك إمكانية استخدامها بسهولة. كما يجب إرفاق بيانات وصفية (Metadata) دقيقة، توضّح محتوى المدونة وبُنيتها، مع الفصل الواضح بين النصوص الخام والبيانات الموسّمة؛ لضمان الشفافية وسهولة المعالجة في التطبيقات المستقبلية.

٦. **التوثيق والبيانات الوصفية:** التوثيق هو عملية مستمرة تبدأ من بداية مشروع التوسيم حتى إطلاق المدونة الموسّمة وتشغيلها، ويساعد التوثيق في الحفاظ على تفاصيل العمل في المشروع منذ بدايته. ومن المهم توثيق البيانات الوصفية وأي معلومات تتعلق بها غير المحتوى اللُّغوي الرئيس، مثل: مصادر البيانات، واسم المؤلف، وتاريخ الوصول إلى المستند.

٧. **معايير خاصة باللُّغة العربية (ع):** نظرًا لما تتفرَّد به اللغة العربية من خصائص لغوية مميزة على المستويات (الصوتية، والصرفية، والنحوية والدلالية) فمن المناسب اعتماد معايير خاصة تأخذ هذه الخصوصية بعين الاعتبار عند توسيم المدونات اللغوية العربية. وتزداد أهمية ذلك عند التعامل مع اللهجات المحكية، التي تفتقر إلى قواعد معيارية موحدة، مما يستدعي وضع معايير مرنة ومدروسة تُراعي التنوع اللهجي، وتسهم في تمثيل الظواهر اللغوية غير المعيارية بصورة منهجية قابلة للمعالجة الآلية والتحليل اللساني الدقيق.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً



## وفيما يأتي قائمة بأهم معايير توسيم المدونات اللُّغوية الخاصة باللُّغة العربية[٣]:

| المجال | المعيار العام | المعيار التفصيلي | الوصف | المؤشر | آلية القياس |
|---|---|---|---|---|---|
| **أولًا: تحديد الهدف من التوسيم** | ١. تحديد الظاهرة اللُّغوية المُستهدَفة | أ. تحديد عدد الظواهر اللُّغوية المراد دراستُها. | تُحدَّد الظواهر اللُّغوية، مثل: دراسة أقسام الكلام للهجة معينة، دراسة المشاعر في النصوص، البنية اللُّغوية للهجة معينة. وتُوصَف بدقة هذه الظواهر اللُّغوية المراد دراستها، وبناء مدونة موسومة تعكس هذه الظواهر اللُّغوية. | عدد الظواهر المحددة. | مراجعة وثيقة التخطيط. |
| | | ب. تحديد العمق المُستهدَف لدراسة الظاهرة اللُّغوية. | | مستوى التفصيل في وصف الظاهرة. | تحليل وثائق التوسيم النظري. |
| | ٢. تحديد الهدف من توسيم المدونة وإمكانية تحقيقه. | ج. وجود غرض أو أغراض مُحدَّدة من توسيم المدونة. | يُحدَّد الغرض من بناء المدونة الموسَّمة، مثل: بناء مدونة تُستخدَم للتوسيم الآلي لأقسام الكلام باللُّغة العربية، أو بناء نموذج آلي لتصنيف المشاعر. | وجود أهداف واضحة. | مراجعة خطة المشروع. |
| | | د. التحقُّق من تطابق الأهداف مع المخرجات. | تُراجَع الأهداف لبناء المدونة بعد تحديد مواصفاتها؛ للتحقُّق من أن مخرجات المواصفات تُطابق الأهداف. | نسبة التطابق بين الأهداف والمخرجات. | مقارنة بين وثائق التخطيط والتطبيق. |

(٣) تم إدراج مؤشرات أداء متوسطة، لا تتطلب - غالبا - إجراءات مطولة، بحيث يمكن الإفادة منها بيسر وسهولة.

| المجال | المعيار العام | المعيار التفصيلي | الوصف | المؤشر | آلية القياس |
|---|---|---|---|---|---|
| | ٣. تحديد الإطار النظري لوصف الظاهرة اللُّغوية. | أ. تحديد مدى تغطية الإطار النظري للظواهر اللُّغوية المُستهدَفة. | يُختار إطار نظري مناسب - أو مجموعة من الأُطر - لوصف الظواهر اللُّغوية المُستهدَف دراستها، مثل: النظريات النفسية لوصف المشاعر البشرية، أو النظريات اللُّغوية في أقسام الكلام أو النحو للغة أو لهجة معينة. مع التأكيد على شمول هذا الإطار النظري لجميع جوانب الظاهرة اللغوية المراد تحليلها. | نسبة التغطية النظرية. | مطابقة الظواهر مع الإطار النظري. |
| | | ب. تحديد مدى تبنِّي الإطار النظري للظاهرة اللُّغوية. | | مدى الاتساق مع الإطار النظري. | تحليل مقارنة بين التوسيم والتأصيل النظري. |
| | ٤. تحديد المُستخدِمين المُستهدَفِينَ | أ. تحديد نوعية المُستخدِمين المُستهدَفينَ. | تحديد نوعية المُستخدِمين المُستهدَفين للمدونة، مثل: الباحثين، والمدرسين، المهتمين بلهجة معينة، المُتخصِّصين في مجال معين. | عدد الفئات المحددة. | مراجعة دراسة الجدوى والاستخدام. |
| | | ب. قياس تجاوب المُستخدِمين مع المدونة بعد نشرها. | يُقاس عدد المستخدمين ونوعية فئاتهم من حيث التخصص، وهدف الاستخدام، ومستوى التفاعل؛ وذلك لمعرفة مدى التجاوب مع المدونة، وتحديد أثرها في المجتمع العلمي. | نسبة التفاعل. | تحليل سجل الاستخدام. |
| | | ج. تحليل تعليقات ومراجعات المُستخدِمين. | تُجمَع تعليقات المستخدمين ومراجعاتهم بشأن المدونة لتحليلها؛ للتحقُّق من عدم وجود مشكلات تعيق استخدامها لدى الفئات المستهدفة. | عدد المراجعات المحللة. | تحليل نصوص المراجعات. |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعيار التفصيلي | الوصف | المؤشر | آلية القياس |
|---|---|---|---|---|---|
| | | د. متابعة المنشورات التي تستخدم المدونة. | متابعة الأوراق العلمية وغيرها من المنشورات التي تستخدم المدونة؛ للتحقق من تغطيتها وتأثيرها في المجالات المعرفية ذات الصلة. | عدد المنشورات المعتمدة على المدونة. | رصد عبر محركات بحث علمية. |
| | ٥. تحديد الوسوم المُستهدَفة | أ. تحديد عدد الوسوم. | تحدد الوسوم المُستهدَفة وعددها بشكل دقيق. | عدد الوسوم المعرّفة. | مراجعة قوائم الوسوم. |
| | | ب. كتابة وصف واضح للوسوم. | يُكتَب لكلّ وسم مُستهدَف وصفٌ واضحٌ يبين معناه ومجاله. | نسبة الوسوم ذات الوصف الدقيق. | مراجعة دليل التوسيم. |
| | | ج. الاتساق في الوسوم وعدم تداخلها. | التحقق من أن الوسوم مستقلة، ولا تتداخل فيما بينها. | نسبة الوسوم المتمايزة. | تحليل هيكل الوسوم. |
| | | د. تقليل عدد الوسوم للحد الأدنى. | عدم تكرار الوسوم المستهدفة، وتقليلها لأقل عدد ممكن يفي بالغرض من بناء المدونة. | نسبة الالتزام بالتقليل. | مقارنة مع المعيار المعتمد. |
| | ٦. تحديد نطاق المدونة المناسب | أ. تحديد مدى تخصُّص النطاق. | يُحدَّد نطاق أو نطاقات المدونة المستهدفة بدقة، كالمجال القانوني، أو الطبي، أو الإخباري، مع مراعاة تخصُّصها الكافي؛ لضمان شمولها جميع المجالات المستهدفة بشكل متكامل. | درجة تخصص النطاق. | تحليل محتوى المدونة. |
| | | ب. شمولية النطاق لأهداف المدونة. | | نسبة الشمولية. | مطابقة أهداف التوسيم مع النطاق. |
| | ٧. تحديد حجم المدونة المُستهدَف | أ. تحديد عدد المستندات المُستهدَفة للتوسيم. | يُحدَّد الحجم المتوقَّع للمدونة المستهدفة للتوسيم، مع مراعاة أن يكون عدد الكلمات والمستندات كافيًا لتحقيق الغرض. | عدد المستندات المحددة. | مراجعة خطة التجهيز. |
| | | ب. تحديد عدد الجمل المُستهدَفة للتوسيم. | | عدد الجمل المحددة. | إحصاء نصوص المدونة. |
| | | ج. تحديد عدد الكلمات المستهدفة للتوسيم. | | عدد الكلمات المستهدفة. | تحليل إحصائي للنصوص. |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعيار التفصيلي | الوصف | المؤشر | آلية القياس |
|---|---|---|---|---|---|
| | ٨. تنوُّع الظواهر والأوعية اللُّغوية | أ. تحديد أنواع أوعية المدونة وعددها: الأخبار، وسائل التواصل الاجتماعي، الرسائل الجامعية. | التحقُّق من أن المدونة الخام المُستهدَفة للتوسيم تغطي ظواهر لغوية متنوعة، وتتضمن أوعية مختلفة تناسب الهدف، كالأخبار مثلًا. | عدد أنواع الأوعية المستخدمة. | تحليل مصادر البيانات. |
| | | ب. تحديد عدد المناطق الجغرافية لمصادر المدونة. | | عدد المناطق الجغرافية المغطاة. | مراجعة التصنيف الجغرافي للبيانات. |
| | | ج. تحديد نوع المصادر اللُّغوية وعددها (المدونة الخام). | | تنوع المصادر اللغوية. | جرد مصادر المدونة الخام. |
| | ٩. تحديد تكلفة التوسيم | أ. حصر الجوانب المؤثرة في تقدير تكلفة التوسيم | تُحدَّد الميزانية في ضوء عدد من المحددات؛ كحجم المدونة المستهدفة، ونوعها، وتكاليف توظيف المُوسِّمين، والأدوات المستخدمة، وغيرها. | عدد العوامل المحددة في التقدير. | تحليل عناصر التكلفة. |
| | | ب. تحديد التكلفة المادية لتوسيم المدونة. | تُحدَّد الميزانية الخاصة بالمدونة بعد تحديد العوامل المؤثرة في تقديرها، ومن المهم متابعة صرف الميزانية في جميع مراحل بناء المدونة؛ لضمان تحقيق الأهداف بكفاءة. | نسبة التقدير إلى الميزانية الفعلية. | مقارنة خطة التمويل بالتنفيذ. |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعيار التفصيلي | الوصف | المؤشر | آلية القياس |
|---|---|---|---|---|---|
| **ثـانيًا: معـايير إعـداد دليـل التـوسـيـم ومنـصـة التوسيم** | ١. جـودة أمثلة دليل التوسيم ووضوحها | أ. وجـود عـددٍ كافٍ مـن الأمثلـة لـكل وسـم. | أن يشمل دليل التوسيم أمثلة كافية لتوضيح الوسوم خاصة في الحـالات التـي يكـثرفيها الالتبـاس. مع وجـود أمثلة مضادة في هذه الحالات؛ لزيادة الوضوح والدقة. | نسبة الوسـوم التـي تتـضمن أمثلـة. | مراجعة دليـل التوسـيم |
| | | ب. وجود عـددٍ كافٍ من الأمثلة المضادة لكل وسم. | | نسـبة الوسـوم التـي تتضـمن أمثلـة مضـادة. | تدقيق محتوى الدليـل. |
| | | ج. وجود الأمثلـة في سياق واضح. | | نسبـة الأمثـلة المصـاحـبـة بسياق لغوي. | تحليـل سياقي للأمثلة. |
| | | د. وضع أمثلة تُوضِّح حالات الالتباس. | | عـدد حـالات الالتبـاس الموضحة. | مـراجـعـة شـروحـات الوسـوم. |
| | ٢. وضوح دليل التوسيم | أ. إبراز الهدف من التوسيم بشكل واضح في الدليل. | أن يكـون دليـل التـوسـيـم واضحًا، ويـغـطي جـميع جوانب التوسيم؛ بـدايـةً من توضيـح الهـدف من التوسيم، والظاهرة اللُّغوية المستهدفة. وكذلـك وصف شامل للوسوم المستهـدفة وعلاقتها بالظاهرة اللُّغوية والإطار النظري المُستخدَم. | وضـوح الهـدف في مـقـدمـة الدليـل. | تحليل افتتاحية الدليـل. |
| | | ب. تـوضـيـح الظاهـرة اللُّغوية المستهدفة. | | نسبـة الوسـوم المرتبطة بظواهر محـددة. | مطـابـقـة الوسـوم مـع الظواهـر. |
| | | ج. توضيـح الإطـار النظري المُستخدَم لوصـف الظـاهرة اللُّغوية. | | نسبـة التـوافق مـع الإطـار النظـري. | مراجعة الربط النظري. |
| | | د. وجـود وصـف شـامل للوسـوم المستهدفة. | | نسبـة الوسوم ذات الـوصـف الكامـل. | تدقـيـق حقل التعريف لكـل وسـم. |

| المجال | المعيار العام | المعيار التفصيلي | الوصف | المؤشر | آلية القياس |
|---|---|---|---|---|---|
| | ٣. التحسين المستمر لدليل التوسيم | أ. وجود وسيلة لإرسال ملاحظات المُوسِّمين. | أن يكون هناك تحسين مستمر لدليل التوسيم مع الأخذ بعين الاعتبار ملاحظات الموسمين واستفساراتهم حول الدليل، ووجود منهجية لمراجعة دليل التوسيم وتحديثه. ويُقاس التحسين بعدد التحديثات التي نُفِّذت في دليل التوسيم. | توفر قناة ملاحظات نشطة. | فحص أدوات الاتصال في الدليل. |
| | | ب. حصر عدد الملاحظات من المُوسِّمين على دليل التوسيم. | | عدد الملاحظات المستلمة. | مراجعة سجل التفاعلات. |
| | | ج. وجود آلية واضحة لمراجعة دليل التوسيم. | | وجود خطة مراجعة. | مراجعة وثيقة الإجراءات. |
| | | د. حصر عدد التحديثات التي تمَّت على دليل التوسيم. | | عدد الإصدارات المحدثة. | مراجعة سجل الإصدارات. |
| **ثالثًا: معايير إعداد دليل التوسيم ومنصة التوسيم** | ١. تتبُّع الإصدارات المختلفة لدليل التوسيم | أ. وجود رقم تسلسلي لإصدار دليل التوسيم. | وجود آلية لتتبُّع الإصدارات المختلفة لدليل التوسيم، ومنهجية واضحة لترقيم النُّسخ المختلفة وكذلك آلية لاعتماد النسخة ومراجعتها من قِبل المشرفين على عملية التوسيم؛ ضمانًا لجودة المحتوى وتناسقه عبر التحديثات. | نسبة النسخ المرقمة. | مراجعة وثائق النسخ. |
| | | ب. تحديد تاريخ للإصدار لكل نسخة. | | نسبة النسخ بتاريخ إصدار. | فحص بيانات النسخ. |
| | | ج. كتابة ملخص التعديلات على النسخة الجديدة. | | توفر ملخصات تغيير. | تحليل محتوى وثيقة الإصدار. |
| | | د. الاحتفاظ بنُسخ الإصدارات المختلفة. | | نسبة النسخ المحفوظة. | مراجعة المحفوظات. |
| | | هـ. تحديد حالة النسخة الحالية: معتمدة، مقترحة، تحت المراجعة... إلخ. | | وضوح حالة النسخة (معتمدة، مراجعة ...) | تحقق من رأس كل نسخة. |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعيار التفصيلي | الوصف | المؤشر | آلية القياس |
|---|---|---|---|---|---|
| | ٢. قابلية الوصول واستخدام دليل التوسيم | أ. تحسين درجة الوصولية للمعلومات في الدليل. | مدى سهولة الوصول إلى المعلومات في دليل التوسيم، مثل: الوقت المستغرق للوصول لمعلومة ما. وكذلك مدى استخدام المُوسِّمين لدليل التوسيم دون الرجوع إلى المشرفين على عملية التوسيم. | عدد مرات الوصول | التحقق من سجل الوصول |
| | | ب. قياس عدد مرات استخدام دليل التوسيم. | | عدد مرات الوصول | التحقق من سجلّ الوصول |
| | | ج. قياس الوقت المستغرق للوصول إلى المعلومة في دليل التوسيم. | | متوسط زمن البحث عن معلومة | مراجعة زمن البحث عن معلومة |
| | ٣. اختيار منصة وأداة التوسيم المناسبة | أ. اختبار عددٍ كافٍ من منصات التوسيم للاختيار من بينها. | من المهم تجربة عدة منصات للتوسيم في حال كانت ستُستخدم من قِبل عدة أشخاص، مثل: الموسمين، مديري مشروع التوسيم والمشرفين عليه، والمشرفين الفنيين. أما في حال كان التوسيم محدود النطاق، فيُفضَّل استخدام أداة بسيطة تُناسب حجم المشروع. مع الأخذ بعين الاعتبار المعايير الأخرى لتقييم كل أداة، ثُم اختيار الأنسب منها حسب الميزانية المتوفرة. | عدد المنصات المجربة | حصر المنصات المجربة |
| | | ب. وجود عدد كافٍ من المستخدمين الذين اختبروا أداة التوسيم. | | عدد المستخدمين المشاركين في التجريب | حصر المستخدمين المشاركين في التجريب |
| | | ج. اختبار الأداة من عدة أنواع من المستخدمين: المستخدمين المُوسِّمين، ومديري المشروع، والمشرفين الفنيين على الأداة. | | تنوع فئات المستخدمين | تصنيف فئات المستخدمين |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعيار التفصيلي | الوصف | المؤشر | آلية القياس |
|---|---|---|---|---|---|
| | | د. اختبار توافق مخرجات منصة التوسيم مع المخرجات المستهدفة. | | نسبة التوافق | مقارنة مخرجات عملية |
| | | هـ. مناسبة تكلفة منصة التوسيم ونوع الرخصة المعتمدة فيها لطبيعة المشروع وميزانيته. | | نسبة التوافق المالي والتراخيص | مراجعة عقد الاستخدام |
| | ٤. سهولة استخدام منصة التوسيم | أ. وجود دليل لاستخدام أداة التوسيم. | من المهم أن تكون منصة التوسيم سهلةَ الاستخدام مع توفير دليل للمُوسِّمين لاستخدامها. | وجود دليل إرشادي مفعل | فحص ملفات المنصة |
| | | ب. تحديد ومتابعة المشكلات الفنية الشائعة في استخدام منصة التوسيم. | أن تُحدَّد المشكلات الفنية الشائعة، وأن تكون سهلة بحيث لا يضيع وقت الموسمين في استخدام المنصة في غير التوسيم. | عدد المشكلات الشائعة | مراجعة سجل الشكاوى |
| | | ج. حساب الوقت الإضافي في استخدام المنصة غيروقت التوسيم. | | نسبة الوقت الإضافي خارج التوسيم | تحليل جلسات العمل |
| | | د. وجود مجتمع واسع من المستخدمين للمنصة. | من المهم أن تكون منصة التوسيم واسعة الانتشار ولها مجتمع مستخدمين بحيث يسهل الوصول للدعم الفني وحل المشكلات التقنية بسرعة. | عدد المستخدمين النشطين | رصد إحصاءات المنصة |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعيار التفصيلي | الوصف | المؤشر | آلية القياس |
|---|---|---|---|---|---|
| | ٥. صيانة ودعم منصة التوسيم | أ. تقليل قياس الفترة المستغرقة للرد على المشكلات الفنية. | في حال وجود عدد كبير من المُوسِّمين في المشروع، من المهم أن يكون هناك آلية سهلة لرفع المشكلات التقنية والفنية، وألا يستغرق التحديث والصيانة وقتًا طويلًا. | متوسط زمن الاستجابة | سجل الدعم الفني |
| | | ب. تقليل تردُّد عمليات التحديث والصيانة للحد الأدنى. | | عدد التحديثات خلال مدة معينة | مراجعة سجل التحديثات |
| | ٦. قابلية تخصيص المنصة | أ. توفُّر خيارات التخصيص بمنصة التوسيم. | من المهم أن تكون المنصة مرنةً بحيث يمكن تخصيصها لكل مستخدم حسب تفضيلاته، وكذلك أن يكون من السهل تكييفها حسب متطلبات التوسيم المختلفة. وأيضًا، قابلية التعديل على الشفرة البرمجية للمنصة يجعلها أكثر قابليةً للتخصيص. | عدد الخصائص القابلة للتخصيص | مراجعة إعدادات المنصة |
| | | ب. تكيُّف المنصة مع متطلبات التوسيم المختلفة. | | نسبة التوافق مع متطلبات المشروع | مطابقة المواصفات |
| | | ج. سهولة التعديل على الشفرة البرمجية لمنصة التوسيم. | | إمكانية التعديل البرمجي | إجراء اختبار تقني عملي |
| **رابعًا: معايير اختيار المُوسِّمين المناسبين وتدريبهم** | ١. كفاءة الموسمين لعملية التوسيم | أ. تحديد عدد مُوسِّمين مناسب لإنجاز الهدف المطلوب ضمن الإطار الزمني المحدد. | تحديد عدد الموسمين ومؤهلاتهم العلمية وخبراتهم السابقة. إضافةً إلى الخلفية الثقافية واللُّغوية لهم التي تناسب مهمة التوسيم المستهدفة، وكذلك الفئة العمرية والجنس. | عدد الموسمين المناسب | مطابقة أعداد الموسمين مع المهام المطلوبة |
| | | ب. تحديد المؤهلات الأكاديمية المناسبة للموسمين. | | تعريف المؤهلات المناسبة | تدقيق وثائق التوظيف |
| | | ج. تحديد الخبرات المناسبة في التوسيم عند اختيار الموسمين. | | تعريف الخبرات المطلوبة | مراجعة الخبرات السابقة |

| المجال | المعيار العام | المعيار التفصيلي | الوصف | المؤشر | آلية القياس |
|---|---|---|---|---|---|
| | | د. تحديد الخلفية الثقافية واللُّغوية والفئة العمرية للموسمين. | | تعريف الخلفيات المطلوبة | تحليل بيانات ديموغرافية |
| | | هـ.تصميم اختبار كفاءة الموسمين لقبولهم في عملية التوسيم. | يُختبر الموسمون اختبارًا معياريًا يُحدِّد كفاءتهم للتوسيم، بحيث لا ينجح إلا مَن يتجاوز درجة معينة في الاختبار. | وجود اختبار كفاءة | مراجعة إجراءات التوظيف |
| | | و.التحقق من تنوع الخلفيات الثقافية للموسمين. | من المهم أن يكون المُوسِّمون من خلفيات متنوعة تتناسب مع الغرض من التوسيم؛ وذلك لتجنُّب أي انحياز في بيانات التوسيم. يشمل هذا التنوُّع جوانب متعددة، مثل: العمر، والجنس، والخلفية العلمية والثقافية وغيرها، بحسب ما يتطلبه المشروع. | نسبة التنوع الثقافي | تحليل تنوع الفريق |
| | ٢. التدريب المستمر للمُوسِّمين على أداة التوسيم | أ. كفاية فترة التدريب للموسمين للعمل على المشروع. | تدريب الموسمين على دليل التوسيم وأداة التوسيم المستخدمة بشكل مستمر، مع الأخذ بملاحظاتهم حول التدريب وجودته. وكذلك عمل اختبارات قبل التدريب وبعده؛ لقياس مدى تطوُّرهم. | عدد ساعات التدريب لكل موسم | مراجعة سجل التدريب |
| | | ب. تسجيل ملاحظات المتدربين على ورش التدريب وأخذها بعين الاعتبار. | | عدد الملاحظات المسجلة | تحليل نماذج التغذية الراجعة |
| | | ج. تحديد عددٍ كافٍ من ورش العمل. | | عدد الورش المنفذة | سجل أنشطة التدريب |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعيار التفصيلي | الوصف | المؤشر | آلية القياس |
|---|---|---|---|---|---|
| | | د. اختبار كفاءة الموسمين قبل التدريب وبعده. | | نسبة التحسن بعد التدريب | مقارنة نتائج الاختبارين |
| | ٣. تقييم الموسمين بشكل مستمر؛ لضمان الاتساق والجودة في التوسيم. | أ. كفاية عدد مرات التقييم، ومدى تكرارها. | لا يُكتفَى بتدريب الموسمين، بل لا بد من تقييم تقدُّمهم في عملية التوسيم بشكل مستمر كمًّا وكيفًا، وبعدد كاف من مرات التقويم؛ لضمان جودة الأداء. | عدد جولات التقييم | مراجعة خطة ضمان الجودة |
| | | ب. حساب عدد الأمثلة التي عمل عليها كل مُوسِّم. | | عدد الأمثلة لكل فرد | تقارير الإنتاج |
| | | ج. حساب عدد الأمثلة التي اختلف فيها وسم أحد الموسِّمين عن وسوم بقية زملائه. | | نسبة التباين | مقارنة نتائج التوسيم |
| | ٤. رَصْد ملاحظات المُوسِّمين المتعلقة بعملية التوسيم | أ. رَصْد عدد الملاحظات من المُوسِّمين. | في أثناء عملية التوسيم تبرز ملحوظات متعددة للموسّمين، وهذه الملحوظات تعد ذات قيمة عالية في تحسين عملية التوسيم؛ لذا لا بد من رصدها بطريقة منهجية، ومشاركتها مع جميع المُوسِّمين، ووضع الحلول المناسبة لها. | عدد الملاحظات المسجلة | استمارات التغذية الراجعة |
| | | ب. رصد عدد المرات التي تم التجاوب فيها مع الملاحظات. | | نسبة الاستجابة للملاحظات | متابعة إجراءات التحسين |
| | | ج. تطبيق آلية تسمح للمُوسِّمين بمشاركة ملاحظاتهم وحل المشكلات المتعلقة بعملية التوسيم. | | توفر آلية فعالة للتواصل | فحص آليات التواصل الداخلي |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً



| المجال | المعيار العام | المعيار التفصيلي | الوصف | المؤشر | آلية القياس |
|---|---|---|---|---|---|
| | ٥. رصد مكافآت للموسمين والمشاركين | أ. عدالة المكافآت ومناسبتها لطبيعة العمل. | التأكد من وجود مكافآت مادية عادلة للموسمين، وأن تُصرَف في وقتها دون تأخير. | نسبة رضا الموسمين عن المكافآت | استبيان دوري |
| | | ب. مناسبة توقيت صرف المكافآت. | | توفر جدولة المكافآت المالية | مراجعة الإجراءات المالية |
| | ٦. التحقق من توازن عبء العمل بين المُوسِّمين | أ. مناسبة عبء العمل وتوازنه. | التحقق من أن عبء العمل متوازن بين الموسمين؛ بهدف تعزيز جودة الأداء وتجنُّب الإجهاد الذي قد يُؤثر سلبًا على مستوى التوسيم. | تفاوت التوزيع بين الموسمين | تحليل أداء الموسمين |
| | | ب. الأخذ بملاحظات الموسمين حول عبء العمل. | | نسبة الأخذ بالملاحظات | تقارير الإدارة الوسطى |
| | ٧. المحافظة على بقاء العاملين في المشروع. | أ. قياس معدل بقاء العاملين في المشروع. | التحقُّق من عدم تسرُّب العاملين في مشروع التوسيم؛ وخصوصًا مَن لديهم خبرة كبيرة. | نسبة الاستمرارية | تتبع سجل التعاقدات |
| | | ب. قياس رضا العاملين في المشروع. | | مؤشر رضا الموظفين | استبيان دوري |
| **خامسًا: معايير التحقق من جودة التوسيم** | ١. وجود عدة موسمين للنصوص نفسها | أ. كفاية عدد المُوسِّمين للمقارنة. | أن يُوسَم النص (المثال) الواحد من أكثر من شخص؛ للمحافظة على الموثوقية والجودة. وأيضًا لقياس مدى الاختلاف بين الموسمين من المهم عدم تغييرهم عبر المراحل المختلفة. | حجم العينة المقارنة | تحليل منهجي لبيانات الجودة |
| | | ب. اتساق مجموعة الموسمين وثباتها بين الأمثلة. | | نسبة التوافق في الأداء | تحليل نتائج الموسمين المتكررة |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعيار التفصيلي | الوصف | المؤشر | آلية القياس |
|---|---|---|---|---|---|
| | ٢. قياس الاتفاق بين الموسمين | أ. قياس معامل الاتفاق بين الموسمين، مثل معامل كوهين كابا. | تُقاس نسبة الاتفاق بين الموسمين كنسبة مئوية، وباستخدام معاييرإحصائية مثل معامل كوهين كابا، إضافةً إلى وجود آلية لحل حالات الاختلاف بين الموسمين؛ مثلًا: بالتصويت بالأغلبية، أو مراجعتها من قِبل خبير. | قياس الاتفاق بمقياس معروف مثل كابا | استخدام مقياس كابا أو مشابه |
| | | ب. قياس نسبة الاتفاق بين الموسمين. | | نسبة الاتفاق | استخدام مقياس كابا أو مقياس مشابه |
| | | ج. وَضْع آلية لحل حالات الاختلاف. | | توفر آلية فعالة | مراجعة الوثائق التشغيلية |
| | ٣. تنقيح التوسيم ومراجعته | أ. حصر عدد مرات التنقيح لكل مثال. | يُراجَع التوسيم بشكل دوري بناءً على ملاحظات الموسِّمين وحالات الاختلاف بينهم. كما تُحدَّد فترة للمراجعة بعد كل مرحلة أو جزء من التوسيم، وتُراجَع حالات الاختلاف، وقياس الاتفاق بين الموسِّمين، والتأكد من اتساق التوسيم وجودته. | عدد مرات التنقيح | تتبع سجل المراجعات |
| | | ب. تدوين ملاحظات المُوسِّمين ومراجعتها. | | عدد الملاحظات المراجعة | مراجعة استمارات التقييم |
| | | ج. حَصْر عدد مرات تكرار المراجعة الشاملة. | | عدد جولات المراجعة الشاملة | مراجعة جدول العمل |
| | | د. تحديد مدى عمق المراجعة الشاملة. | | مستوى العمق في المراجعة | تحليل تقارير الجودة |
| | | هـ. تحديد عدد العوامل التي يتم مراجعتها. | | عدد العوامل قيد المراجعة | فحص قوائم التحقق المستخدمة |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعيار التفصيلي | الوصف | المؤشر | آلية القياس |
|---|---|---|---|---|---|
| | ٤. وَضْع منهجية واضحة لحل حالات الاختلاف بين المُوسِّمين | أ. تحديد الوقت المستغرق لحل حالات الاختلاف. | تُوضع منهجية واضحة لمعالجة الحالات التي يختلف فيها الموسمون، مع حساب الوقت المستغرق لذلك، وكذلك كفاءتها في الوصول لحل مقبول. تتضمن هذه المنهجية، على سبيل المثال: إحالة الحالات إلى خبير متخصص، أو تنظيم جلسات نقاش بين الموسِّمين والمشرفين على مشروع التوسيم؛ لضمان التوصل إلى حلول مقبولة وموثوقة. | متوسط وقت الحل | تتبع سجلات الحل الزمني |
| | | ب. تحديد كفاءة المنهجية في حل حالات الاختلاف وحدودها. | | مستوى فعالية منهجية الحل | تحليل نتائج تطبيق المنهجية |
| | ٥. وضع حلقة لتداول الملاحظات بين المراجعين والمُوسِّمين | أ. حصر عدد الملاحظات وتكرارها. | وجود آلية وضع حلقة بين المراجعين والموسمين للتعامل مع الملاحظات؛ لتحسين جودة التوسيم باستمرار. | عدد الملاحظات المتكررة | مراجعة نماذج الملاحظات |
| | | ب. تحديد عدد مرات أخذ الملاحظات في الاعتبار. | | نسبة الاستجابة للملاحظات | تحليل أثر الملاحظات على التحديثات |
| | ٦. إجراء فحوصات جودة دورية لحل مشكلات التوسيم | أ. تحديد مدى تكرار الفحوصات. | إجراء فحوصات للجودة بانتظام؛ لتحديد أي مشكلة في عملية التوسيم، وحلها، وإيجاد آلية لجمع الملاحظات حول جودة التوسيم وأخذها في الاعتبار؛ لتعزيز الأداء وتحسين النتائج باستمرار. | عدد جولات الفحص | مراجعة خطة التدقيق |
| | | ب. تحديد عدد المشكلات التي تمَّ حلها. | | عدد المشكلات المعالَجة | مقارنة تقارير ما قبل/بعد الفحص |
| | | ج. تحديد الوقت المستغرق لحل المشكلات. | | متوسط زمن الحل | متابعة زمن الاستجابة |
| | | د. تحديد عدد المشكلات المُعلَّقة. | | عدد المشكلات المعلقة | فحص سجل المهام المفتوحة |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعيار التفصيلي | الوصف | المؤشر | آلية القياس |
|---|---|---|---|---|---|
| | | هـ. تحسين الجودة وفقًا للملاحظات. | | نسبة التحسين الفعلي | تحليل نسخ التوسيم قبل/ بعد |
| | ٧. اتساق التوسيم | أ. قياس مدى درجة اتساق التوسيم لنفس الأمثلة. | التحقُّق من اتساق التوسيم في الأجزاء المختلفة من المدونة، مثل أن تكون الكلمة نفسها لها وسم مختلف ضمن سياق آخر في المدونة. وكذلك معالجة حالات الاختلاف إنْ وجدت. | نسبة الاتساق الداخلي | تحليل التكرارات |
| | | ب. حَصْر عدد الوسوم للمثال نفسه ضمن سياقات مختلفة. | | متوسط عدد الوسوم للمثال الواحد | تحليل الوسوم المتعددة للأمثلة |
| | | ج. وضع آلية لحل الاختلاف بين الوسوم غير المتسقة. | | توفر آلية معالجة التناقض | مراجعة وثائق الحل والاعتماد |
| | ٨. تقسيم التوسيم على دفعات | أ. تحديد حجم مناسب للعينة لكل دفعة. | يُقسَّم التوسيم على دفعات صغيرة، بحيث تُراجع كل دفعة بشكل مستمر، ويُزاد من حجم العينة لكل دفعة تدريجيًّا بعد التحقُّق من زيادة الجودة للمستوى المطلوب. ويجب أن يكون وقت العمل على الدفعة مناسبًا بحيث يُستكمَل بعد مراجعة الدفعات السابقة من المشرفين بوقت كافٍ. | متوسط حجم العينة لكل دفعة | مراجعة خطة التوسيم المرحلي |
| | | ب. تحديد عدد المراجعات اللازمة لكل دفعة. | | عدد المراجعات لكل دفعة | تحليل جداول المراجعة المرحلية |
| | | ج. رَصْد التغيرات التي تتم بعد كل دفعة. | | عدد التغيرات الملحوظة | مقارنة التوسيم قبل وبعد الدفعات |
| | | د. مناسبة الوقت المستغرق للانتهاء من الدفعة. | | متوسط الزمن لكل دفعة | تتبع جداول الإنجاز اليومية |

| المجال | المعيار العام | المعيار التفصيلي | الوصف | المؤشر | آلية القياس |
|---|---|---|---|---|---|
| سادسًا: معايير تخزين المدونة الموسّمة | ١. استخدام ترميز معياري | أ. التحقق من اتساق الترميز المُستخدَم. | تخزين المدونة الموسّمة بترميز معياري مثل UTF-8 وصيغة ملفات معيارية مثل xml بحيث تكون متوافقةً مع مختلف المنصات وأنظمة التشغيل. واتساق ذلك في مختلف مراحل التوسيم وعدم تغييره. | نسبة اتساق الترميز | مراجعة تقنية لعينة من الملفات |
| | | توافق الترميز المُستخدَم مع أنظمة الترميز المعيارية. | | مستوى التوافق المعياري | اختبار توافق الملفات مع المعايير الدولية |
| | ٢. تحسين درجة المقروئية للمدونة | ب. تسجيل الوقت المستغرق للوصول لبيانات المدونة. | تخزين المدونة بشكل يسهل للمُستخدِم تصفُّحها وليس فقط للآلة. وكذلك وجود أدوات مثل واجهة المستخدم الرسومية؛ للتعامل مع المدونة بشكل سهل. ومن ذلك تقسيم المدونة بشكل منطقي بحسب الهدف منها، مثل تقسيمها إلى فترات زمنية مختلفة أو تقسيمها حسب الموضوعات. | متوسط زمن الوصول | تتبع استخدام النظام |
| | | ج. تدوين ملاحظات المُستخدِمين حول المقروئية. | | عدد الملاحظات المسجلة | استبيانات وتجميع التغذية الراجعة |
| | | د. حساب وتقليل الوقت اللازم للمُستخدِمين للوصول إلى المعلومة. | | نسبة التحسن في وقت الوصول | مقارنة نتائج القياس الدوري |
| | | هـ. تصميم أدوات مناسبة لواجهة المُستخدِم الرسومية. | | عدد الأدوات المتاحة | مراجعة واجهة المستخدم وتقارير التطوير |
| | ٣. فصل البيانات الخام عن التوسيم | أ. التحقُّق من كفاءة فصل الوسوم عن البيانات الخام. | التحقُّق من فصل البيانات الخام عن التوسيم بطريقة لا تؤثر على البيانات؛ لضمان سلامة البيانات الأساسية وعدم التأثير عليها. | نسبة البيانات المفصولة بدقة | مراجعة العينة يدويًا أو آليا |
| | | ب. التحقُّق من سلامة البيانات بعد الفصل. | | نسبة سلامة البيانات | اختبار التحقق من سلامة البيانات |

| المجال | المعيار العام | المعيار التفصيلي | الوصف | المؤشر | آلية القياس |
|---|---|---|---|---|---|
| | ٤. عمل نسخ احتياطية بشكل دوري | أ. تكرار النسخ الاحتياطية بشكل دوري. | أن تُنسخ المدونة بشكل دوري كنسخة احتياطية؛ للحفاظ على سلامة البيانات، مع التحقق من فعالية هذه النسخ في الحفاظ على سلامتها. | عدد النسخ المنفذة | مراجعة سجل النسخ الاحتياطي |
| | | ب. التحقُّق من استعادة البيانات في النسخ الاحتياطية. | | نسبة الاسترجاع الناجح | اختبار عشوائي لاستعادة البيانات |
| | ٥. تطبيق آلية لتسلسل الإصدارات المختلفة، وتوثيق الاختلافات بينها | أ. توثيق الإصدارات المختلفة للمدونة. | تطبيق آلية تراعي ضبط تسلسل الإصدارات المختلفة منعًا للتداخل بينها، مع توثيق التغييرات لكل نسخة جديدة، والاختلافات بينها وبين النسخ السابقة. | عدد الإصدارات الموثقة | مراجعة سجل الإصدارات |
| | | ب. ترقيم الإصدارات المختلفة بأرقام تسلسلية. | | توفّر تسلسل لترقيم الإصدارات | مراجعة سجل الإصدارات |
| | | ج. توثيق التغييرات على كل نسخة جديدة. | | مستوى التوثيق للاختلافات | مقارنة الإصدارات وتحليل التعديلات |
| | ٦. المحافظة على أمن بيانات المدونة | أ. تطبيق معايير الأمن السيبراني. | الحفاظ على بيانات المدونة من التعديل والتخريب. | تطبيق معايير الأمن المحلية والعالمية | التأكد من تطبيق المعايير |
| | | ب. حماية بيانات المدونة من الوصول غير المشروع. | تطبيق الطرق التقنية التي تُنظِّم عملية الدخول إلى بيانات المدونة، مثل: استعمال اسم المُستخدِم وكلمة المرور، أو طُرق التحقُّق الثنائي للمُستخدِمين. | نسبة التشفير الناجح | اختبار أمني لعينة من الملفات |
| | | ج. تطبيق أنظمة حماية البيانات، وضمان خصوصيتها. | التأكد من حفظ بيانات المدونة وتخزينها وفقًا لسياسات الخصوصية وحماية البيانات الشخصية. | توفر حماية فاعلة | اختبارات الاختراق وتقارير الأمان |

| المجال | المعيار العام | المعيار التفصيلي | الوصف | المؤشر | آلية القياس |
|---|---|---|---|---|---|
| **سابعًا: التوثيق والبيانات الوصفية** | ١. تحديث معلومات المدونة وتوثيق عملية التحديث | أ. وجود مستند مُحدَّث يصف كل مرحلة بالتفصيل، مع تاريخ الإصدار. | التحقق من أن التوثيق مكتمل ويغطي جميع مراحل بناء المدونة، بما فيها دليل التوسيم والتعديلات التي نفذت. | توفر المستندات المحدثة | مراجعة وثائق المشروع |
| | | ب. تسجيل عمليات التحديث للمدونة وتوثيقها. | تُحدَّث مستندات التوثيق وفقًا لملاحظات المستخدِمين، وكذلك أي عملية تحديث نُفذت على المدونة، مع توثيق نسبة الأخطاء أو الالتباسات التي تُرصد بعد كل عملية تحديث، ومدى انخفاضها أو ثباتها بمرور الزمن. | معدل التوثيق الزمني | مراجعة سجل التحديثات |
| | | ج. التأكد من سلامة المعلومات المُضافة أو المُعدَّلة في كل إصدار جديد للتوثيق. | | نسبة السلامة في التحديث | مراجعة داخلية لمحتوى التوثيق |
| | ٢. وضوح مستندات التوثيق، والتحقق من سهولة الوصول إليها | أ. اعتماد لغة واضحة وسلسة في مستندات التوثيق. | التحقق من أن مستندات التوثيق واضحة وسهلة للقراءة، ومراجعة ملاحظات المستخدمين حول ذلك مع توفير وسائل لتدوين ملاحظاتهم. | مستوى وضوح اللغة | تحليل نصي باستخدام معايير الوضوح |
| | | ب. توفير قنوات أو أدوات تتيح للمُستخدِمين الإبلاغ عن أي صعوبات في فهم أو تطبيق محتوى التوثيق. | | عدد قنوات التغذية الراجعة | مراجعة خصائص المنصة |
| | | ج. توفير وسائل بحث أو فهرسة واضحة داخل منصات التوثيق. | التحقق من أن مستندات التوثيق يسهل الوصول لها من المستخدمين، وتوثيق ملاحظاتهم وأخذها بعين الاعتبار. | توفر أدوات البحث | اختبار قابلية البحث والفهرسة |
| | | د. توثيق ملحوظات المُستخدِمين الواردة حول الوصولية. | | عدد الملاحظات الموثقة | مراجعة قاعدة بيانات الملاحظات |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعيار التفصيلي | الوصف | المؤشر | آلية القياس |
|---|---|---|---|---|---|
| | ٣. التحقق من اكتمال البيانات الوصفية وصحتها واتساقها ودقة حفظها | أ. مدى تغطية البيانات الوصفية وشموليتها: تاريخ الجمع، تاريخ النشر، طول المستند، المؤلِّف... إلخ. | إضافة البيانات الوصفية كاملةً لكل مستند في المدونة، مع مراعاة دقتها وصحتها. | نسبة التغطية | مراجعة آلية لنماذج البيانات |
| | | ب. التحقق من صحة البيانات الوصفية المضافة، عبر مقارنتها مع المصدر أو مراجعتها من قِبل المشرفين على المشروع. | | مستوى دقة البيانات الوصفية | مراجعة ثنائية أو آلية مقارنة |
| | | ج. ضمان اتساق البيانات الوصفية. | مراعاة أن تكون البيانات الوصفية متسقةً بين مختلف المستندات والوثائق في المدونة، وأن تكون موافقة للمعايير. | نسبة الاتساق | اختبار تطابق بنيوي للبيانات |
| | | د. الالتزام بالمعايير اللازمة في حفظ البيانات الوصفية. | | توافق الحفظ مع المعايير | تدقيق في لهيكلية البيانات |
| | | هـ. ضمان حفظ البيانات الوصفية في ملف أو قاعدة بيانات منفصلة عن المحتوى الأصلي. | مراعاة أن تكون البيانات الوصفية مفصولةً عن بيانات المدونة دون الإخلال بسلامة البيانات وصحتها. | تحقق الفصل بين المحتوى والبيانات | فحص بنية التخزين |
| | | و. التأكد من عدم إتلاف محتوى المدونة عند إضافة أو تحديث البيانات الوصفية. | | نسبة المحتوى السليم | مراجعة بعدية للمحتوى |

| المجال | المعيار العام | المعيار التفصيلي | الوصف | المؤشر | آلية القياس |
|---|---|---|---|---|---|
| **ثـامـنًا: معـاييـر خاصة باللُّغة العربيـة** | ١. استخـدام ترميز معيـاري لـلأحـرف العربيـة | أ. استخـدام تـرميز مُوحَّـد عند تخزين البيانـات العربيـة. | استخدام ترميز معياري مثل UTF-8 للغة العربية، بحيث يكون دقيقًا، ويحافظ على النص بشكل سليم. | نسبـة الملفات المرمّـزة بشكـل معيـاري | تـدقيـق بنـية الملفات |
| | | ب. التحقق دوريًا من سـلامة النـص بعـد أي عمليـة تحويـل أو استيراد. | | عـدد الأخطاء المكتشفـة بعد التحويـل | مراجعة التقارير التقنية |
| | ٢. استخـدام أنظمة النسـخ الحـرفي | أ. استخـدام نظـام نسـخ حـرفي يغطي جميـع الحـروف والعلامـات العربيـة. | في حال عدم استخدام ترميز خـاص بالعربيـة أو عالمي مثـل UTF-8، يُستـخدَم نظام للنسخ الحـرفي مـثل Buckwalter بحيث يغطي كل الحـروف العربيـة وعلامات التشكيـل، مع وجـود آلية للتفريـق بـين النـص العـربي المكتـوب بنظـام النـسخ الحـرفي والنـص بلغـة أخـرى. | اكتمـال تغطيـة النظـام | مقارنـة الجـداول المرجعيـة |
| | | ب. وَضْـع آلـيـة للتفـريـق بيـن النـص العـربي المنسـوخ حرفيًا والنـص الأصلي بلغـة أخـرى. | | وجـود آليـة فعالـة للتفريـق | مراجعـة الوثائق والأدوات |
| | ٣. التحقُّـق مـن دقـة عمليـة التشكـيـل واتسـاقها | أ. اسـتخدام مصـادر موثوقـة أو أدوات تدقيـق آلي عنـد إضافـة التشـكيل. | التحقُّـق مـن أن التشكـيـل صحـيح ومتسـق ولا سيمـا التشـكيل الاختيـاري. | نسـبـة الاعتمـاد عـلـى مصـادر موثوقـة | مراجعـة مصـدر التشـكيل |
| | | ب. اعتمـاد سـياسة واضحة للتشكيل، والحـفـاظ على اتسـاقهـا عبـر النصـوص. | | نسـبـة الاتسـاق في التشـكيل | فـحص العينـات آليًـا وبشـريًا |

| المجال | المعيار العام | المعيار التفصيلي | الوصف | المؤشر | آلية القياس |
|---|---|---|---|---|---|
| | ٤. مراعاة القواعد الإملائية | أ. الاستناد إلى دليل إملائي مُحدَّد. | تعريف قواعد إملائية واضحة، واتساق التوسيم في ذلك، والتحقق من صحته. ويشمل ذلك التحقق من التباين في كتابة الكلمات، مثل: "سعودي" مقابل "سعودى"، ولا سيما في حال توسيم اللهجات. | وجود دليل إملائي موحد | فحص العينات آليًا وبشريًا- ومراجعة الدليل. |
| | | ب. الالتزام بالكتابة الصحيحة للكلمات الشائعة، والتفريق بين التهجئة الفصيحة واللهجات. | | نسبة الالتزام بالقواعد | تدقيق إملائي دوري |
| | | ج. وَضْع معيار واضح لكيفية كتابة الاختصارات. | | وجود دليل للاختصارات | مراجعة الدليل |
| | ٥. مراعاة دقة التحليل الصرفي واتساقه | أ. توحيد الوسوم والأكواد المستخدمة في التوسيم الصرفي. | وسم الجذور الثلاثية والرباعية، وتقطيع الكلمات صرفيًا بشكل متسق وصحيح. ووضع معايير واضحة ومحددة للمحافظة على الاتساق والصحة، وكذلك تحديد مدى التفصيل في التحليل الصرفي (مثل: تحديد السوابق واللواحق). | وجود دليل توسيم صرفي موحد | مراجعة الدليل |
| | | ب. تحديد مستوى التفصيل المطلوب. | | وجود مستوى التفصيل | مراجعة لغوية |
| | | ج. تنفيذ مراجعة دقيقة من مختص لغوي للتحليل الصرفي. | | تحقق مراجعة المختص | تحليل سجلات الأخطاء والمراجعة |
| | | د. صحة التقطيع الصرفي. | | تحقق مراجعة التقطيع الصرفي | سجلات الأخطاء والمراجعة |
| | | هـ. التحقق من الاتساق في السوابق واللواحق والمقحمات. | | عدد الأخطاء الموثقة | سجلات الأخطاء والمراجعة |

| المجال | المعيار العام | المعيار التفصيلي | الوصف | المؤشر | آلية القياس |
|---|---|---|---|---|---|
| | ٦. وجود آلية ومعايير واضحة لتقطيع النص إلى جمل وكلمات | أ. الاستناد إلى علامات الترقيم والمسافات والروابط اللُّغوية لتجزئة النص إلى جمل. | وجود آلية ومعايير واضحة لتقطيع النص إلى جمل وكلمات، مثل تقطيعها بناءً على علامات الترقيم والفراغات. وكذلك آلية لتقطيع الكلمات إلى وحداتها الصرفية. | نسبة التقطيع الصحيح | مراجعة النصوص المقطعة |
| | | ب. تقسيم الجمل إلى كلمات وفق سياسة معيّنة. | | وجود سياسة تقسيم | مراجعة السياسة |
| | | ج. توضيح ما يُعَدُّ "وحدة صرفية" عند التعامل مع السوابق واللواحق (مثل: الواو، الفاء،... إلخ)؛ لضمان الاتساق. | | وجود قواعد واضحة لاتساق التقسيم | مراجعة القواعد |
| | | د. إجراء عمليات تدقيق يدوية أو آلية بعد التقطيع. | | مراجعة التقطيع يدويا وآليا | اختبار الحالات الخاصة |
| | ٧. وَضْع ضوابط دقيقة للتعامل مع اللهجات | أ. تعريف نطاق اللهجة جغرافيًا. | تُعرَّف اللهجة المُستهدَفة حسب النطاق الجغرافي: دولة، مدينة، منطقة. ويُراعَى أن يكون هذا التعريف متسقًا. وكذلك يُراعَى وجود نصوص تتضمن تناوبًا لغويًا (بحيث يُبدِّل المتحدِّث بين عدة لهجات في النص نفسه). | عدد اللهجات المحددة | مراجعة وثائق المشروع |
| | | ب. الالتزام بالأسلوب الإملائي نفسه داخل اللهجة الواحدة. | | وجود قواعد مكتوبة | تدقيق وثائق العمل |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعيار التفصيلي | الوصف | المؤشر | آلية القياس |
|---|---|---|---|---|---|
| | | ج. وَضْـع سياسة واضحة لحالات يُبـدَّل فيها المُتحـدِّث بين لهجات أو بين العربية الفصحى ولهجة محكية ضمن النص الواحد. | | توفّـر سياسة التحوّل اللغوي | مراجعة سياسات التوسيم |
| | ٨. مراعاة دقة التحليل النحوي واتساقه | أ. اعتماد قواعد نحوية مرجعية واضحة. | توضع معايير وتعريفات واضحة للتوسيم النحوي، وعمق المعلومات النحوية المستهدفة (مثل: الاكتفاء بتعيين الوظيفة النحوية أم تضمين البنية الشجرية الكاملة للجملة). | توفر التحليل النحوي | اختبار النتائج بمراجع نحوية |
| | | ب. تحديد مدى تفصيل المعلومات. | | نسبة المطابقة | مراجعة بشرية |
| | ٩. وَضْع قواعد واضحة ومُحدَّدة للتفريغ الصوتي | أ. وجود قواعد مكتوبة للتفريغ الإملائي للصوتيات. | تُوضَع قواعد واضحة للتفريغ الصوتي ولا سيما للهجات المحكية، بحيث تكون متسقةً، وتغطي جميع الوحدات الصوتية في اللُغة. | وجود معايير مكتوبة | مراجعة الوثائق |
| | | ب. استخدام رموز صوتية موحَّدة. | | توحيد الرموز الصوتية | مراجعة الدليل |
| | | ج. التأكد من تمثيل جميع الفونيمات الرئيسة في اللهجة. | | التحقق من دقة التمثيل | مراجعة بشرية/آلية |

| المجال | المعيار العام | المعيار التفصيلي | الوصف | المؤشر | آلية القياس |
|---|---|---|---|---|---|
| | ١٠. دَعْم منصة التوسيم للعربية | أ. إتاحة الكتابة باتجاهين (من اليمين إلى اليسار، ومن اليسار إلى اليمين) في منصة التوسيم. | التحقق من أن منصة التوسيم تدعم اللُّغة العربية والكتابة باتجاهين، وتسمح بالتخصيص والتغيير بشكل مرن؛ لمراعاة خصائص اللُّغة العربية. | نسبة التوافق مع العربية | اختبار عملي على النظام |
| | | ب. السماح بتعديل الواجهة أو آليات التوسيم؛ لتتناسب مع خصوصيات اللُّغة العربية. | | توفّر الدعم | مراجعة واجهة المنصة |

## المراجع :


- الثبيتي، عبد المحسن (٢٠٢١). دليل توسيم البيانات اللُّغوية. منشورات مدينة الملك عبد العزيز للعلوم والتقنية.
- العصيمي، صالح فهد. (Ed.). (٢٠١٥). المدونات اللُّغوية العربية:: بناؤها وطرائق الإفادة منها.
- الفيفي، عبد الله يحيى. (٢٠٢٣). وسم المدونات اللُّغوية: المفهوم والمجالات. سلسلة العلوم الإنسانية والاجتماعية، ١(٣٨)، ص ٢٧٥-٣١٦.

- Atkins, S., Clear, J., & Ostler, N. (1992). Corpus design criteria. *Literary and Linguistic Computing, 7*(1), 1–16. https://doi.org/10.1093/llc/7.1.1
- Belinkov, Y., Habash, N., Kilgarriff, A., Roth, R., & Suchomel, V. (2013). arTenTen: a new, vast corpus for Arabic. *In Proceedings of WACL* (p. 20).
- Garside, R., & Leech, G. N. (Eds.). (1997). *Corpus annotation: Linguistic information from computer text corpora*. Longman.
- Ide, N. (1998). Corpus Encoding Standard: SGML Guidelines for Encoding Linguistic Corpora. *In LREC* (pp. 463–470).
- Ide, N., & Pustejovsky, J. (Eds.). (2017). *Handbook of linguistic annotation*. Springer Netherlands.
- Ide, N., & Suderman, K. (2014). The linguistic annotation framework: a standard for annotation interchange and merging. *Language Resources and Evaluation,* 48, 395–418.
- Jarrar, M., Khalilia, M., & Ghanem, S. (2022). Wojood: Nested Arabic named entity corpus and recognition using BERT. *In Proceedings of the Thirteenth Language Resources and Evaluation Conference* (pp. 3626–3636).

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً


- Lai, V. D., Ngo, N. T., Veyseh, A. P. B., Man, H., Dernoncourt, F., Bui, T., & Nguyen, T. H. (2023). ChatGPT beyond English: *Towards a comprehensive evaluation of large language models in multilingual learning*. arXiv preprint arXiv:2304.05613.
- Maamouri, M., Bies, A., Buckwalter, T., & Mekki, W. (2004). The Penn Arabic Treebank: Building a large-scale annotated Arabic corpus. *In NEMLAR conference on Arabic language resources and tools* (Vol. 27, pp. 466–467).
- Marcus, M., Santorini, B., & Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics, 19*(2).
- Pustejovsky, J., & Stubbs, A. (2012). *Natural language annotation for machine learning.* O'Reilly Media, Incorporated.

# الفصل السابع

## معايير دعم اللُّغة العربية في التطبيقات الحاسوبية

- ملخص
- مقدمة
- مدخل تعريفي إلى دعم اللُّغة العربية في التطبيقات الحاسوبية
- مُسوِّغات بناء المعايير لدعم اللُّغة العربية في التطبيقات الحاسوبية
- الدراسات السابقة
- نطاق المعايير وحدودها
- المراجع



هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

## ملخص

يستعرض هذا الفصل عددًا من معايير التطبيقات الحاسوبية المُوجَّهة للمُستخدِم العربي، وتشمل: معالجة المدخلات، والمخرجات، والنصوص باللغة العربية. وتتضمَّن المعايير مجموعةً من الضوابط التي ينبغي مراعاتها في برمجة التطبيقات الحاسوبية لدعم اللُّغة العربية، إضافةً إلى المعايير ذات العلاقة بتصميم مواقع الشبكة العالمية الإنترنت. وقد صُنِّفت المعايير إلى عشر فئات تتضمن: التصميم، والتشغيل، ومعايير الأمان والأداء، وتجربة المُستخدِم، وغيرها من الموضوعات الرئيسة التي تتكون منها صناعة التطبيقات والمنصات الحاسوبية المختلفة.

## مقدمة

في عصر التقنية الرقمية، أصبحت التطبيقات الحاسوبية جزءًا رئيسًا من حياتنا اليومية، سواء كانت تُستخدَم في الهواتف الذكية، أو الحواسيب، أو الأجهزة اللوحية. وتتنوع هذه التطبيقات بين الترفيهية، والتعليمية، والإنتاجية، وحتى تلك التي تساعد في إدارة الحياة اليومية. ومع تزايد الاعتماد على التطبيقات، برزت الحاجة إلى وضع معايير لتطويرها وضمان توافقها مع اللُّغة العربية وثقافتها، بما يخدم المُستخدِم في العالم العربي.

وتُعَدُّ معايير التطبيقات أداة حيوية لضمان الأداء الفعَّال والسلس لهذه البرامج. فهي تسهم في تحقيق مستويات عالية من الأمان، والحماية من الثغرات الأمنية التي قد تُعرِّض بيانات المستخدمين الشخصية للخطر. إضافةً إلى ذلك، تُسهم المعايير في تحسين تجربة المُستخدِم عن طريق ضمان سهولة الاستخدام، والتوافق مع مختلف الأجهزة وأنظمة التشغيل. فضلًا عن أنَّ وضع معايير لتطوير التطبيقات يسهم في تعزيز الابتكار؛ إذ يُقدِّم للمُطوِّرين إطارًا مُوحَّدًا يُمكنهم عن طريقه التركيز على الإبداع وتحسين الوظائف بدلًا من الانشغال بحل مشكلات التوافق والأداء. وبالنسبة للشركات، فإن الالتزام بهذه المعايير يُعزِّز من سمعتها وثقة المستخدمين في

منتجاتها؛ مما يفتح لها آفاقًا جديدةً للنمو والانتشار في الأسواق. وبالتالي، تقوم معايير التطبيقات بعمل محوريٍّ في تحقيق التوازن بين الأمان، والأداء، والجودة؛ وهو ما يضمن للمستخدمين تجربةً سلسةً وآمنةً، وللمُطوِّرين بيئةً مُحفِّزة على الابتكار والإبداع.

## مدخل تعريفي

تهدف هذه الوثيقة إلى وَضْع معايير تُؤخَذ بعين الاعتبار عند تطوير التطبيقات الحاسوبية Computer Applications أو مواقع الشبكة العالمية (الإنترنت) Websites التي تستخدم اللُّغةَ العربية لغةً طبيعيةً، سواء بشكل مستقل أو ضمن لغات أخرى. وتغطي هذه المعايير كذلك تطبيقات الأجهزة الذكية. آخذةً بعين الاعتبار التعامل مع المحتوى العربي بدايةً من كتابة شفرة البرمجيات مرورًا بتصميم واجهة المُستخدِم بكل مكوناتها النصية أو الرسومية أو الوسائط المتعددة، وسواء كان ذلك في مرحلة المدخلات أو المخرجات. إضافةً إلى ذلك، تستعرض هذه الوثيقة المعايير الخاصة بتطبيقات الذكاء الاصطناعي ونماذجه في دعم اللغة العربية؛ نظرًا للتطور الكبير الذي شهده هذا المجال في الآونة الأخيرة.

## تعريف مختصر بفئات المعايير

### تندرج المعايير تحت مجموعة من الفئات، يمكن تلخيصها في الآتي

- **كتابة الشفرة (الكود):** وهي تسبق إطلاق التطبيق أي: عندما يبدأ المُطوِّرون كتابة شفرة التطبيق. وهنا ينبغي أن تُؤخذ سمات اللُّغة العربية بعين الاعتبار منذ بداية هذه المرحلة.
- **سمات اللُّغة العربية:** للغة العربية خصائصها بدايةً من المستوى الصوتي، فالصرفي والنحوي، ثم الدلالي، ويُضاف إلى ذلك وسائل التواصل المختلفة (الصوتية والكتابية). ومن المهم أَخْذ هذه الخصائص بعين الاعتبار.

- **اللُّغة العربيـة ضمـن لغـات أخـرى:** كثيرًا مـا تكـون التطبيقـات متاحـةً بأكـثر مـن لغة. وفي الغالـب يكـون التطبيـق بلغـة غـير عربيـة ثـم يترجـم إلى العربيـة؛ وهـو مـا يُسـبِّب عجـزًا وقصـورًا في المحتـوى العـربي أو الوظائـف المرتبطـة باللُّغـة العربيـة. لهـذا، مـن المهـم أن تكـون اللُّغـة العربيـة حاضـرةً مـع اللُّغـات الأخـرى عنـد تطويـر التطبيقـات.
- **فئـات عمـر المُسـتخدِم:** تختلـف اللُّغـة المُسـتخدَمة باختـلاف الفئـات العمريـة للمُسـتخدِمين. فالأطفـال ليسـوا كالكبـار مـن حيـث مسـتواهم اللُّغـوي ووسـائل التفاعـل مـع التطبيـق؛ لـذا مـن المهـم أن تؤخـذ كل فئـة مـن هـذه الفئـات بعـين الاعتبـار، وتُحـدَّد الفئـة العمريـة المسـتهدفة مـن كل تطبيـق.
- **مسـتوى تعليـم المُسـتخدِم:** مثلما تختلـف الفئـات المسـتهدفة مـن الناحيـة العمرية، فإنها تختلـف أيضًـا مـن الناحيـة التعليميـة، ممـا يـؤدي إلى تبايـن في المسـتوى اللُّغـوي داخـل التطبيـق؛ فلغـة التطبيقـات المُوجَّهـة للعامـة غـير تلـك المُوجَّهـة للباحثـين والمهـن المختصـة. لذلـك، ينبغـي مراعـاة مسـتوى تعليـم كل فئة عنـد تطويـر التطبيق، وتحديـد الفئـة المسـتهدفة عنـد إطلاقـه؛ لضمـان تحقيـق أفضـل تجربـة اسـتخدام.
- **التوطـين:** يختلـف التوطـين عـن الترجمـة؛ فالترجمـة تشـير إلى تحويـل الكلمـات والتعابيـر مـن اللُّغـة الأصليـة إلى لغـة أخـرى دون تغيـير المعـنى، أمـا التوطـين فهـو عمليـة شـاملة تتضمَّـن التركـيز علـى عناصـر أخـرى غـير مقترنـة بالنـص إلى جانـب الأبعـاد الثقافيـة والعناصـر اللُّغويـة مـن أجـل تطويـع ومواءمـة هُويـة العلامـة التجاريـة بمـا ينسـجم مـع الموقـع الجغـرافي والسـياق الثقـافي للجمهـور المسـتهدف. ولا بـد مـن التأكـد مـن تحديـد النطـاق الكلي لأهـداف توطـين الموقـع الشـبكي، وتهيئـة المُطوِّريـن والمُترجِمـين وإعدادهـم للتعامـل مـع أي تحديـات يمكـن أن تواجههـم عنـد التطبيـق.

## مُسوِّغات بناء المعايير

شـهدت السـنوات الأخـيرة تناميًـا متسـارعًا في اسـتخدام التطبيقـات الحاسـوبية في مختلـف جوانـب الحيـاة، سـواء عـبر الأجهـزة المكتبيـة أو الهواتـف الذكيـة. وفي السـياق العـربي

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً



تحديـدًا، تـبرز حاجـة ملحّـة إلى وضـع معايـير تضمـن دعمًـا لغويًّـا وثقافيًّـا متكامـلًا، يجعـل هـذه التطبيقـات أدوات موثوقـة وفعَّالـة للمُسـتخدِم العـربي. وتنبـع هـذه الحاجـة مـن جملـة مُسـوِّغات رئيسـة، لعـل مـن أهمها: التنـوُّع الواسـع الـذي تتسـم بـه اللُّغـة العربيـة ولهجاتها، ومـا يرتبـط بذلـك مـن تحديـات في توحيـد آليـات الكتابـة والواجهـة. كمـا أنَّ تزايـد الاعتمـاد علـى التقنيـة في دول المنطقـة يُحفِّـز المُطوِّريـن والشـركات علـى الاهتمـام بجوانـب، مثـل: الاتجـاه النـصي (الكتابـة مـن اليمـين إلى اليسـار)، ودقـة الترجمـة، وصحـة الصـرف والنحـو، فضـلًا عـن ضـرورة الالـتزام بالممارسـات العالميـة في الترميـز والتعامـل مـع النصـوص، مثـل: الترميـز المُوحَّـد (UNICODE).

في الوقـت ذاتـه، يتطلَّـب هـذا التوجُّـه ضمـان اتسـاق المعايـير في مختلـف مراحـل تصميـم التطبيـق وتطويـره، بحيـث لا ينشـأ عنهـا تشـتُّت أو تفـاوت في الجـودة أو قابليـة الاسـتخدام. فمـن دون وجـود أُطـر واضحـة يُحتَكَـم إليهـا، قـد يُجمِّـع مطـوِّرون نصوصًـا عربيـة بخطـوط أو أنظمـة ترميـز متباينـة، أو يُهملـون -في فـترات لاحقـة- تنفيـذ جوانـب تصحيـح الأخطـاء اللُّغويـة وتمكـين واجهـات عربيـة سـليمة؛ الأمـر الـذي ينعكـس سـلبًا على تجربـة المُسـتخدِم وقدرتـه علـى الثقـة في المنتـج التقـني المحلـي. ولعـل جانـب التكامـل مـع المكتبـات والأدوات العالميـة في مجـال الـذكاء الاصطناعـي ومعالجـة اللُّغـة الطبيعيـة يُعَـدُّ مؤشّـرًا آخـر علـى أهميـة وجـود معايـير وطنيـة أو إقليميـة رصينـة، تضمـن أن تطبيقـات التحليـل والتقسـيم النـصي والتعـرُّف علـى الكيانـات وغيرهـا تعمـل علـى نحـو متوافـق مـع الواقـع اللُّغـوي والثقـافي للعـرب.

ومـن هـذه المُسـوِّغات تنطلـق المبـادئ الرئيسـة الـتي اعتمدتهـا هـذه الوثيقـة إذ تقـوم أولًا علـى مبـدأ **الشـمول**، بمعـنى توفـير معايـير تغطـي سلسـلةً واسـعةً مـن السـيناريوهات مـن مجـرد دعـم النـص العـربي في مخرجـات بسـيطة، إلى توطـين التطبيقـات المعقَّـدة للـذكاء الاصطناعـي. وتلـتزم الوثيقـة كذلـك بمبـدأ **المرونـة والتحديـث** فاللُّغـة العربيـة ليسـت حبيسـة قوالـب ثابتـة، إنمـا تتطـوَّر وتتفـرَّع بحسـب اللهجـات والمجـالات المعرفيـة؛ ممـا يفـرض أن تظـل المعايـير قابلـةً للتعديـل مـع ظهـور تحديـات جديـدة. وفي الإطـار نفسـه، يجـري

**التوافـق مـع الممارسـات العالميـة** في الترميـز والتصميـم وتوحيـد البيانـات الأمـر الـذي يُرسِّـخ إمكانيـة الاسـتفادة مـن تجـارب مماثلـة، ويُضاعـف قابليـة التكامـل مـع المنصـات الدوليـة.

إضافـةً إلى ذلـك، تؤكّـد هـذه المبـادئ علـى مراعـاة **تجربـة المُسـتخدِم العـربي وثقافتـه**، حيـث تؤكِّـد هـذه المبـادئ مراعـاة **البُعـد الثقـافي** في التصميـم والتطويـر، إذ إنَّ اللُّغـة والثقافـة عنصـران لا ينفصـلان، فـإلى جانـب الالـتزام بتقديـم محتـوى عربي سـليم مـن الناحيـة اللُّغويـة، ينبغـي أَخْـذ الخصوصيـات الثقافيـة بعـين الاعتبـار، بمـا في ذلـك دعـم التقويمـات المحليـة (مثـل: التقويـم الهجـري)، وأسـماء الأشـهر المتداولـة في كل بلـد (مثـل: ينايـر مقابـل كانـون الثـاني)، وتبـنِّي الأشـكال الشـائعة اجتماعيًّـا للأرقـام وللصياغـات الزمنيـة. فالتطبيـق الـذي يحـترم ثقافـة المُسـتخدِم، ويتبـنَّى صيـغ التواريـخ والأسـاليب الحسـابية المتداولـة في مجتمعـه؛ يمنـح الشـعور بالأُلفـة والارتيـاح، ويُعـزِّز تفاعلـه مـع محتوياتـه وخدماتـه. وفي السـياق ذاتـه، يسـتهدف مبـدأ **ضمـان جـودة المعالجـة اللُّغويـة** رفـع دقـة النصـوص العربيـة وتصويبهـا نحويًّـا وصرفيًّـا، حـتى في التفاصيـل التقنيـة الدقيقـة، مثـل: أسـاليب التشـكيل ورمـوز اللُّغـة؛ لمـا لذلـك مـن أثـر في موثوقيـة التطبيـق. أخيـرًا، ينعكـس هـذا النهـج المُنظِّـم علـى تيسـير الوصـول للأسـواق وزيـادة التنافسـيَّة؛ إذ تصبـح التطبيقـات المُوجَّهـة إلى المُسـتخدِم العـربي أكـثر إقناعًـا وحضـورًا، وتُنافـس بدورهـا حلـولًا عالميـة قـد لا تراعـي خصوصيـات اللُّغـة أو الثقافـة العربيـة بالقـدر ذاتـه.

بهـذا تـبرز المعايـير المُبيَّنـة في هـذه الوثيقـة إطـارًا مرجعيًّـا يجمـع بـين الشـمول والمرونـة والمواءمـة العالميـة، مُتخـذةً مـن احتياجـات المُسـتخدِم العـربي ومكانـة لغتـه محـورًا ترتكـز إليـه جميـع التفاصيـل الفنيـة واللُّغويـة، بـدءًا مـن كتابـة الشـفرة البرمجيـة ووصـولًا إلى تجربـة المُسـتخدِم النهائيـة.

## الممارسات والتطبيقات السابقة

لتحليـل واقـع دعـم اللُّغـة العربيـة في التطبيقـات الحاسـوبية، يمكـن تتبُّـع جـذور هـذا الدعـم منـذ اخـتراع الآلـة الكاتبـة وانتشـارها تجاريًّـا في القـرن التاسـع عشـر، ثـم تطويـر أول

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً



آلـة كاتبـة تدعـم العربيـة وغيرهـا مـن اللُّغـات الـتي تسـتخدم أنظمـة كتابـة غـير اللاتينيـة. ومـع ظهـور الحاسـوب في منتصـف القـرن العشـرين، تنـوَّع توزيـع الحـروف في لوحـة المفاتيـح (أو مـا يُعـرَف بتخطيـط لوحـة المفاتيـح). كذلـك مـع انتشـار الحاسـوب، اختلفـت طُـرق ترميـز وتخزيـن النصـوص والحـروف في أنظمـة التشـغيل المختلفـة؛ ممـا أسـفر عـن مشـكلات في التوافـق بـين التطبيقـات والأنظمـة المختلفـة. لذلـك، بـرزت الحاجـة إلى إعـداد معايـير لتوزيـع أحـرف الكتابـة وترميـز النصـوص، وغيرهـا مـن الضوابـط التقنيـة الـتي تسـهّل تطويـر تطبيقـات ومعـدات حاسـوبية تدعـم اللغـة العربيـة، وتضمـن توافقهـا مـع غيرهـا مـن الأنظمـة البرمجيـة.

وقـد تطـوَّر دعـم اللُّغـة العربيـة في الحاسـوب تدريجيًّـا منـذ أواخـر السـبعينيات مـن القـرن المـاضي. حيـث كانـت التحديـات تتمثَّـل في إيجـاد نُظـم ترميـز للحـروف (character encoding) تدعـم اللُّغـة العربيـة، وكذلـك تصميـم خوارزميـات برمجيـة تدعـم الكتابـة العربيـة بحيـث يمكـن تغيـير شـكل الحـرف حسـب موقعـه مـن الكلمـة، فضـلًا عـن التعامـل مـع الحـركات (الفتحـة، والضمـة، والكسـرة، والشـدة، والسـكون) كأحـرف خاصـة.

ولحـق ذلـك في الثمانينـات ظهـور أنظمـة تشـغيل تدعـم اللُّغـة العربيـة، مثـل: نظـام دوس ٣,٣ مـن شـركة مايكروسـوفت MS-DOS، ثـم تطويـر شـركة العالميـة لحاسـوب صخـر AH-200 في عـام ١٩٨٥م المبـني علـى نظـام تشـغيل بدعـم كامـل للغـة العربيـة. وعليـه، ظهـرت الحاجـة إلى وجـود معايـير مرجعيـة لدعـم اللُّغـة العربيـة وفقًـا لمتطلبـات التطبيقـات حينهـا. ولعـل مـن أوائـل المعايـير الـتي صُمِّمـت هـو معيـار ASMO 449 الصـادر مـن المنظمـة العربيـة للتوحيـد القيـاسي والمواصفـات (آزمـو ASMO) حينهـا (دُمجـت حاليًّـا تحـت منظمـات أخـرى)، وتـم تسـجيله لاحقًـا في المنظمـة الدوليـة للمعايـير (الأيـزو) تحـت رمـز ISO 9036، وهـو معيـار لترميـز الحـروف العربيـة باسـتخدام نظـام ٧ بـت، ويدعـم الأحـرف العربيـة فقـط دون اللاتينيـة. وتـم تطويـر معيـار آخـر لاحقًـا مـن قبـل منظمـة الآزمـو تحـت رمـز ASMO 708، يدعـم كلًّا مـن الأحـرف العربيـة واللاتينيـة (ثنـائي الاتجـاه)، وتـم تسـجيله لاحقًـا في الأيـزو تحـت رمـز ISO 8859-6.

مع مرور الزمن، تطوَّرت أنظمة الترميز في دعم اللُّغة العربية حتى ظهور معيار الترميز المُوحَّد Unicode كنسخة أولى في عام ١٩٩١م، وهو نظام ترميز يدعم معظم أنظمة الكتابة المختلفة في العالم. ويُعَدُّ ظهور Unicode نقطة تحول مهمة في دعم اللُّغة العربية واللُّغات الأخرى في التطبيقات الحاسوبية. وقد أُنشئ اتحاد يونيكود (Unicode Consortium) بوصفه هيئة معايير غير ربحية تهدف إلى توحيد ترميز الحروف من مختلف أنظمة الكتابة في العالم؛ مما يسهل عملية التبادل والتوافق بين الأنظمة والتطبيقات المختلفة.

قبل ظهور Unicode، كانت هناك العديد من نُظم الترميز المختلفة التي تعاني من محدودية في عدد الأحرف المدعومة وتعارضات بينها؛ مما أدى إلى صعوبات في معالجة النصوص متعددة اللُّغات. جاءت الحاجة إلى نظام ترميز شامل يدعم جميع اللُّغات، وهنا برز دور Unicode في توفير حل لهذه المشكلات من خلال تخصيص رمز فريد لكل حرف أو رمز في أي لغة.

يدعم Unicode معظم أنظمة الكتابة في العالم، بما في ذلك اللُّغة العربية بجميع حروفها وتشكيلاتها. وهذا يشمل دعم الحركات العربية، وتغيير شكل الحرف حسب موقعه في الكلمة، إضافةً إلى دعم الاتجاه النصي من اليمين إلى اليسار، وهو أمر جوهري للغة العربية واللُّغات المشابهة. كما يتيح Unicode تمثيل نصوص متعددة اللُّغات في تطبيق واحد دون تعقيدات كبيرة؛ مما يُيسّر على المطورين إنشاء تطبيقات تدعم عدة لغات، بما فيها العربية، دون الحاجة إلى إجراء تغييرات جوهرية في بنية التطبيق.

مع مرور السنوات، تطوَّر معيار Unicode ليشمل المزيد من الأحرف واللُّغات والرموز، حيث وصل في نسخته الأحدث إلى ترميز أكثر من ١٥٤,٠٠٠ حرف ورمز. كما أصبح Unicode الأساسَ الذي تعتمد عليه معظم أنظمة التشغيل الحديثة، مثل: ينـدوز (Windows)، وماك أو إس (macOS)، ولينكس (Linux)، ولغات البرمجة مثل: جافا (Java) وسي شارب (C#)؛ وهو ما يعزز من إمكانية التوافق والتكامل بين الأنظمة المختلفة.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً



وأسـهم اعتمـاد Unicode في تسـهيل تطويـر تطبيقـات تدعـم اللُّغـات المتعددة، بمـا في ذلك العربيـة، دون الحاجـة إلى التعامـل مـع تعقيـدات نظـم الترميـز القديمـة. كمـا أسـهم في تحسـين تجربـة المُسـتخدِم، وتعزيـز القـدرة علـى التواصـل والتبـادل الثقافي عـبر الحـدود اللُّغويـة.

وإضافـةً إلى الجهـود المبذولـة في تطويـر معايـير الترميـز وتوزيـع الحـروف علـى لوحـات المفاتيـح، ظهـرت بعـض الدراسـات والمعايـير الخاصـة بواجهـات المُسـتخدِم الرسـومية (User Interface – UI) وتجربـة المُسـتخدِم (User Experience – UX) الـتي تضـع في الحسـبان أهـمَّ المعايـير للتفاعـل الحيـوي بـين الحاسـوب والمُسـتخدِم بشـكل عـام دون تخصيـص لغـة معينـة. هـذه المعايـير تهـدف إلى تحسـين التفاعـل وسـهولة الاسـتخدام للمُسـتخدِمين مـن مختلـف الخلفيـات الثقافيـة واللُّغويـة.

وقـد شـهدت العقـود الأخـيرة تطـوُّرًا ملحوظًـا في دعـم اللُّغـات المختلفـة خاصـة اللُّغـة العربيـة في التطبيقـات الحاسـوبية مـن قِبـل شـركات تقنيـة المعلومـات العالميـة. فعلـى سـبيل المثـال: خصَّصـت شـركة أوراكل Oracle في وثيقتهـا "Oracle Application Server Reports Services" فصـلًا كامـلًا لدعـم العولمـة والدعـم ثنـائي الاتجـاه، ممـا يعكـس (Pamadi ,2004) الأهميـة المتزايـدة لتوفـير معايـير لتطويـر تطبيقـات تدعـم اللُّغـات المتعـددة، بمـا في ذلـك العربيـة. وتُسـلِّط الوثيقـةُ الضـوءَ علـى أهميـة تصميـم تطبيقـات يمكـن أن تُنشـر في بلـدان مختلفـة، مـع مراعـاة مجموعـات الأحـرف واتجـاه النـص. حيـث يتيـح دعـم العولمـة اسـتخدام مجموعـات أحـرف دوليـة، وعـرض البيانـات وفقًـا للمعايـير اللُّغويـة والإقليميـة المناسـبة، إضافـةً إلى إمكانيـة ترجمـة النصـوص في واجهـة المسـتخدِم لتلائـم الثقافـات المختلفـة. أمـا بالنسـبة للدعـم ثنـائي الاتجـاه، فتُوضـح الوثيقـة كيفيـة تصميـم تطبيقـات تدعـم اللُّغـات الـتي تُكتَـب مـن اليمـين إلى اليسـار، مثـل العربيـة، عـن طريـق التحكُّـم في اتجـاه التخطيـط وترتيـب القـراءة والمحـاذاة. هـذا يضمـن عـرض العناصـر والتفاعـلات بشـكل صحيـح؛ ممـا يعـزز تجربـة المُسـتخدِم العـربي ويجعـل التطبيقـات أكـثر ملاءمـةً وفاعليـةً. كمـا تؤكـد الوثيقة أهميـة اسـتخدام الترميـز المُوحَّـد Unicode، الـذي يسـمح بتمثيـل نصـوص متعـددة اللُّغـات

في تطبيق واحد دون تعقيدات كبيرة. هذا يُسهِّل على المُطوِّرين إنشاء تطبيقات تدعم عدة لغات، بما فيها العربية، دون الحاجة إلى إجراء تغييرات جوهرية في بنية التطبيق.

وإلى جانب جهود شركة أوراكل، قدَّمت شركة IBM في وثيقتها "AIX Globalization" توجيهات شاملة للمُبرمجين ومديري الأنظمة حول كيفية تمكين التطبيقات من دعم التعددية الثقافية في نظام التشغيل AIX. حيث تُركِّز هذه الوثيقة على توفير معلومات حول مبادئ العولمة ومعاييرها، بما في ذلك الموضوعات المتعلقة بالمحليات (Locales)، ومجموعات المحارف (Code Sets)، وطرق الإدخال (Input Methods)، والروتينات الفرعية (Subroutines)، والمحولات (Converters)، وخرائط المحارف (Character Mapping)، والمعلومات الثقافية الخاصة، ونظام الرسائل (Message Facility).(١)

(١) تشير المحارف إلى الرموز الأساسية التي تُستخدم في تكوين النصوص، وتشمل الحروف، الأرقام، والرموز الخاصة مثل "أ"، "c"، و"@"، حيث يُمثَّل كل محرف داخل الحاسوب بقيمة رقمية باستخدام أنظمة ترميز مثل Unicode. وتنتمي هذه المحارف إلى ما يُعرف بـ مجموعات المحارف (Code Sets)، وهي أنظمة تُحدد كيفية تمثيل المحارف رقميًا، مثل ASCII الذي يقتصر على الإنجليزية، و Unicode الذي يدعم لغات متعددة بينها العربية. ولضمان التوافق بين هذه الترميزات، تُستخدم المحوّلات، وهي أدوات تُحوّل البيانات من ترميز إلى آخر، كتحويل النصوص من Windows-1256 إلى UTF-٨ أو تحويل التاريخ من الميلادي إلى الهجري. أما الروتينات الفرعية، فهي أجزاء برمجية تؤدي وظائف محددة داخل البرامج، مثل دالة تقوم بحساب مجموع رقمين أو دالة تُعيد تنسيق التاريخ وفق لغة المستخدم. وتلعب المحليات (Locales) دورًا مهمًا في ضبط إعدادات العرض بناءً على اللغة والمنطقة، مثل ar_SA التي تمثل اللغة العربية في السعودية وتُحدد استخدام التاريخ الهجري والعملة بالريال، أو en_US للغة الإنجليزية في الولايات المتحدة. وفي سياق دعم اللغات غير المدعومة مباشرة بلوحة المفاتيح، تُستخدم طرق الإدخال، مثل أدوات تتيح كتابة اليابانية أو استخدام لوحة مفاتيح عربية افتراضية. وتُعد خرائط المحارف وسيلة لربط المحارف بترميزات مختلفة، فتُستخدم لتصحيح عرض النصوص في البيئات غير المتوافقة، كتحويل "é" من Latin-1 إلى Unicode. وأخيرًا، يُستخدم نظام الرسائل لإدارة النصوص الظاهرة للمستخدم داخل التطبيقات، حيث يُعتمد على مفاتيح رسائل مثل MSG_INVALID_PASSWORD، ويتم ترجمتها تلقائيًا حسب اللغة المختارة، كأن تظهر "كلمة المرور غير صحيحة" بالعربية، أو "Mot de passe invalide" بالفرنسية، مما يسهل تعريب التطبيقات وتعدد لغاتها.

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً



تهدف هذه الجهود إلى إنشاء نظام لا يعتمد على افتراضات أو اعتمادات مسبقة على لغة معينة أو ثقافة محددة. حيث يتم الحصول على جميع المعلومات المتعلقة بالثقافة واللُّغة أثناء وقت التشغيل؛ مما يسمح للتطبيقات بمعالجة البيانات وعرضها بشكل صحيح وفقًا للمعايير الثقافية واللُّغوية للمستخدمين. ويشمل ذلك عدم الاعتماد على مجموعات محارف محددة، أو تصنيفات المحارف، أو قواعد المقارنة والترتيب، أو تنسيقات الأرقام والعملات، أو تنسيقات التاريخ والوقت، أو لغة نصوص الرسائل.

وعن طريق تقديم هذه الأدوات والمعايير، أسهمت IBM في تمكين المُطوِّرين من إنشاء تطبيقات تدعم لغات متعددة، بما في ذلك اللُّغة العربية، وتتوافق مع متطلبات الثقافات المختلفة. فضلًا عن ذلك، عرضت الوثيقة التحديات المرتبطة بالاتجاه الثنائي للنصوص (Bidirectionality) وتشكيل الحروف (Character Shaping)، وهو أمرٌ جوهريٌّ للغات التي تُكتب من اليمين إلى اليسار مثل العربية؛ مما يسهم في تعزيز تجربة المُستخدِم العربي، ويضمن عرض المحتوى والتفاعلات بشكل صحيح وفعَّال.

هذه الجهود من قِبل شركات تقنية المعلومات العالمية، مثل: أوراكل وIBM، تعكس الحاجة المستمرة إلى تطوير معايير تراعي خصوصية اللُّغة العربية وثقافتها في تصميم التطبيقات الحاسوبية. وعلى الرغم من وجود بعض الوثائق التي تعرض جوانب محددة من دعم اللُّغة العربية؛ إلا أنه لا يزال هناك حاجة إلى وثيقة شاملة تجمع هذه المعايير، وتُقدِّم إرشاداتٍ واضحةً ومُحدَّثة للمُطوِّرين؛ بما يسهم في بناء تطبيقات متوافقة ومتجانسة، تلبي احتياجات المستخدمين العرب، وتعزز من تجربتهم التقنية.

ونظرًا لارتباط اللُّغة بالثقافة؛ تبرز الحاجة الملحة لتطوير معايير تراعي اللُّغة وثقافتها في تصميم التقنيات والتطبيقات الحاسوبية، التي صار استعمالها وتصفُّحها جزءًا رئيسًا في كل تفاصيل حياتنا اليومية. وعلى الرغم من وجود بعض الوثائق التي تختص ببعض المعايير ذات العلاقة بجوانب دعم التطبيقات للغة العربية، مثل وثائق معايير الترميز الخاصة باللُّغة؛ إلا أن الحاجة لا تزال قائمةً لمزيد من الجهود المُنظّمة في هذا المجال.

وفي إطار السعي نحو تطوير المعايير التي تدعم الثقافات المختلفة في تقنية المعلومات، برز المعيار الدولي ISO/IEC 15897:2011 بعنوان "تقنية المعلومات- واجهات المستخدم- إجراءات تسجيل العناصر الثقافية" (ISO, 2011) (Procedures for the registration of cultural elements). هذا المعيار -الصادر عن المنظمة الدولية للمعايير (ISO) واللجنة الدولية الكهروتقنية (IEC)- يركز على إجراءات تسجيل المواصفات الثقافية المتعلقة بواجهات المستخدم في مجال تقنية المعلومات.

ويهدف المعيار إلى توفير إطار عمل لتسجيل العناصر الثقافية، بما في ذلك المواصفات النصية الثقافية، وملفات الإعدادات المحلية (Locales)، وخرائط المحارف (Charmaps) وذلك وفقًا لمعايير .POSIX ومن خلال هذا النظام الموحَّد للتسجيل، تتوحَّد المعايير الثقافية؛ مما يسهم في تسهيل تطوير تطبيقات متوافقة مع الثقافات واللُّغات المختلفة، بما في ذلك اللُّغة العربية.

ويوفِّر المعيارُ معلوماتٍ مُفصَّلةً عن كيفية تعريف وتسجيل العناصر الثقافية بشكل يمكن معالجته آليًّا، مثل: استخدام لغة الترميز القابلة للتوسعة (XML)، ولغة الترميز القياسية العامة (SGML)، وهذا يتيح للمُطوِّرين والمُنتِجين في قطاع تقنية المعلومات الوصول إلى قاعدة بيانات مُوحَّدة للمواصفات الثقافية؛ مما يُيسِّر عملية تطوير التطبيقات وتكييفها مع المتطلبات الثقافية للمُستخدِمين في مختلف البلدان.

وعلى الرغم من أن المعيار ISO/IEC 15897:2011 يُقدِّم إطارًا مهمًّا لتسجيل العناصر الثقافية، إلا أنه يتداخل جزئيًّا مع مشروعات أخرى، مثل: المستودع المشترك للبيانات المحلية (Common Locale Data Repository CLDR) الذي يُقدِّمه اتحاد يونيكود unicode، أو الترميز المُوحَّد(2024) UNICODE . وهذا المشروع يُوفِّر بيانات محلية بصيغة XML للاستخدام في التطبيقات الحاسوبية، ويهدف إلى توحيد البيانات المحلية وتحديثها لدعم المزيد من اللُّغات والثقافات، بما في ذلك اللُّغة العربية.

كما يُقدِّم CLDR لبنات البناء الرئيسة للبرمجيات لدعم لغات العالم، حيث يحتوي على بيانات محلية تغطي مجموعةً واسعةً من العناصر، مثل:

- **أنماط تنسيق وتحليل** التواريخ، والأوقات، والمناطق الزمنية، والأرقام، والقيم النقدية، ووحدات القياس.
- **ترجمات أسماء** اللُّغات، والنصوص، والبلدان والمناطق، والعملات، والعصور، والأشهر، وأيام الأسبوع، والفترات الزمنية، والمدن، ووحدات الزمن.
- **معلومات عن اللُّغات والنصوص:** تشمل الحروف المستخدمة، وحالات الجمع، والجنس، والقوائم، والقواعد المتعلقة بالترتيب والتصنيف، واتجاه الكتابة، وقواعد تحويل النصوص، وقواعد تقسيم النصوص إلى وحدات (حروف، كلمات، جمل).
- **معلومات عن البلدان:** مثل: استخدام اللُّغة، ومعلومات العملات، وتفضيلات التقويم، ومعايير الأسبوع.
- **إعدادات الصلاحية:** تعريفات وإعدادات الصلاحية للغات، والنصوص، والمناطق، والامتدادات.

كما تعتمد العديد من الشركات والمؤسسات الكبرى على CLDR في منتجاتها وخدماتها، **بما في ذلك:**

- **أبل:** Apple في أنظمة tvOS، watchOS، iOS، macOS، والعديد من التطبيقات.
- **قوقل:** Google في خدمات البحث، متصفح كروم، نظام أندرويد، والعديد من الخدمات الأخرى.
- **آي بي إم:** IBM في منتجات، مثل: DB٢، Tivoli، WebSphere، Lotus.
- **ميتا:** Meta في منصات الفيسبوك، والماسنجر، والواتساب.

- **مايكروسـوفت**:Microsoft في أنظمـة الويـنـدوز، والأوفيـس، وبيئـة التطويـر فيجـوال سـتوديو .Visual Studio

وغيرهـا مـن الشـركات والمؤسسـات الـتي تسـتخدم CLDR لتعزيـز دعـم اللُّغـات والثقافـات في تطبيقاتهـا.

بالنسـبة للغـة العربيـة، يسـهم CLDR بشـكل كبـير في توفـير البيانـات والمعايـير اللازمـة؛ لضمـان عـرض المعلومـات وتنسـيقها بشـكل صحيـح، بمـا في ذلـك الأرقـام، والتواريـخ، والعمـلات، واتجـاه النصـوص؛ ممـا يضمـن تجربـة اسـتخدام متكاملـة تتوافـق مـع الخصوصيـات اللغويـة والثقافيـة للمُسـتخدِم العـربي.

وعلـى الرغـم مـن أن CLDR يتداخـل جزئيًّـا مـع المعيـار ISO/IEC 15897 فيمـا يتعلـق بالبيانـات المحليـة، إلا أنـه يُعتـبر أكـثر شـمولًا وتحديثًـا، حيـث يغطـي أكـثر مـن ٤٠٠ لغـة، ويخضـع لتحديثـات منتظمـة تواكـب التطـورات اللغويـة والثقافيـة المسـتمرة، ممـا يجعلـه مرجعًـا معتمـدًا لتوفـير تجربـة اسـتخدام متعـددة اللغـات.

وتعكـس هـذه المبـادرات والجهـود المسـتمرة الحاجـة الملحـة إلى وجـود معايـير موحـدة وشـاملة تدعـم اللُّغـة العربيـة في التطبيقـات الحاسـوبية. ورغـم التقـدُّم المُحـرَز عـن طريـق معايـير، مثـل: ISO/IEC 15897، ومشـاريع، مثـل: CLDR؛ إلا أن الحاجـة لا تـزال قائمـةً لوثيقـة مرجعيـة شـاملة توحِّـد هـذه المسـاعي، وتقـدِّم توجيهـات واضحـة ومحدَّثـة للمطوريـن، تراعـي الخصوصيـات الثقافيـة واللغويـة للمسـتخدم العـربي. مـن شـأن ذلـك أن يسـهم في تطويـر تطبيقـات أكـثر توافقًـا وفعاليـةً، تسـتجيب بشـكل أفضـل لاحتياجـات الناطقـين باللغـة العربيـة.

## نطاق المعايير وحدودها:

تسـعى هـذه الوثيقـة إلى تأسـيس إطـار معيـاري شـامل يأخـذ في الحسـبان متطلبـات تطويـر التطبيقـات الحاسـوبية ومواقـع الإنترنـت الـتي تعتمـد اللغـة العربيـة لغـةً أساسـيةً أو ضمـن لغـات متعـددة. وتشـمل هـذه المعايـير أيضًـا تطبيقـات الأجهـزة الذكيـة، مـع التركيـز

على التعامل المتقن مع المحتوى العربي في جميع مراحله، سواء من خلال كتابة الشفرات البرمجية أو تصميم واجهات المستخدِم التي تضم النصوص، والرسوم، والوسائط المتعددة، في كل من مدخلات النظام ومخرجاته. كما تتناول الوثيقة المعايير المرتبطة بتطبيقات الذكاء الاصطناعي ونماذجها التي تدعم اللغة العربية؛ نظرًا لما شهدته هذه التقنيات من تطور ملحوظ في الآونة الأخيرة.

## قائمة المعايير:

تعرض الوثيقة الآتية المعايير الخاصة التي يمكن أن تُسهم في دعم العربية في التطبيقات الحاسوبية، وتتمثل في الآتي:

| المجال | المعيار العام | المعيار التفصيلي | الوصف | المؤشر | آلية القياس |
|---|---|---|---|---|---|
| **أولًا: المعايير البرمجية** | ١. مراعاة المعايير العالمية في كتابة الرموز البرمجية (الكود) | أ. حفظ الحروف العربية بصيغة الرمز المُوحَّد Unicode. | التعامل مع الحروف العربية وفقًا لأفضل الممارسات العالمية. | نسبة الملفات الداعمة لـ Unicode | تدقيق الملفات البرمجية |
| | | ب. استخدام النمط المتعاقب Cascading Style Sheets؛ ليسهل التحكم في خلفيات النصوص وألوانها. | مراعاة المعايير العالمية فيما يتعلق بحقوق الأشخاص ذوي الاحتياجات الخاصة. | نسبة استخدام CSS | فحص التصميم البرمجي |
| | | ج. استخدام الحروف الإنجليزية في تسمية الملفات وعناوين قواعد البيانات؛ لتتماشى مع شفرة البرمجة. | | نسبة الأسماء المعيارية | تدقيق أسماء الملفات وقواعد البيانات |

| المجال | المعيار العام | المعيار التفصيلي | الوصف | المؤشر | آلية القياس |
|---|---|---|---|---|---|
| | | د. إمكانية تغيير ألوان الخلفيات والحروف. | | عدد خيارات التخصيص | مراجعة واجهة المستخدم |
| | | هـ. إمكانية دمج محتوى التطبيق في برمجيات أخرى كالناطق الآلي، والتعرف على الكلام. | | نسبة الدمج الناجح | اختبارات الدمج البرمجي |
| | | و. الكتابة بالحروف اللاتينية في رموز لغات البرمجة المختلفة. | تُكتب لغة البرمجة بحروف عالمية يمكن لأي مُبرمج على هذه اللُّغات فهمها والتعامل معها؛ وذلك تفاديًا لأي خلل برمجي bug بسبب عدم توافق لغة البرمجة مع لغات البرنامج الأخرى ذات العلاقة بالتطبيق أو الموقع المستهدف. | نسبة الرموز المكتوبة بحروف لاتينية | مراجعة الرموز |
| ثانيًا: المكتوب العربي | | أ. إخفاء الروابط التشعبية على النص العربي. | عدم عرض نصوص بحروف غير عربية إلا للضرورة كما في رموز العناصر الكيميائية، بما في ذلك الروابط التشعبية (تكون مُضمَّنة في النص)، وفي حال الضرورة تُوضع الحروف غير العربية بين هلاليين (example). | نسبة ا لنصو ص الخالية من الروابط | فحص واجهات المستخدم |
| | | ب. وجود ترجمة للنصوص غير العربية عند ورودها. | | نسبة النصوص المترجمة | مراجعة الترجمة في التطبيق |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعيار التفصيلي | الوصف | المؤشر | آلية القياس |
|---|---|---|---|---|---|
| | | ج. وَضْع الحروف غير العربية بين قوسين هلاليين. | | نسبة الحروف الموضوعة بين هلالين | تدقيق العرض النصي |
| | ١. مراعاة سمات الكتابة العربية وخصائصها | د. كتابة الحروف والأرقام العربية بصيغة رقمية على شكل نص. | استخدام الحروف العربية كنص وليس كصورة، يُسهِّل الاستفادة من المحتوى في التطبيق نفسه أو خارج التطبيق، مثل نسخ النص ومعالجته رقميًّا. | نسبة الحروف والأرقام المكتوبة كنص | تدقيق تقني لملفات النص |
| | | هـ. إمكانية تظليل النصوص ونسخها. | | وجود خاصية التظليل للنسخ | اختبار واجهة المستخدم |
| | | و. استخدام اللُّغة العربية الفصيحة. | سلامة المحتوى من الأخطاء اللُّغوية بجميع أنواعها، وأن يكون بلغة عربية فصيحة يفهمها مُتحدِّثو جميع اللهجات العربية. | نسبة النصوص بالفصحى | مراجعة لغوية |
| | | ز. وضوح مخارج الأصوات. | | نسبة وضوح المخارج | اختبار نطق آلي وبشري |
| | | ح. سلامة النصوص المكتوبة من الأخطاء الإملائية. | | نسبة النصوص السليمة | تدقيق إملائي دوري |
| | | ط. مراعاة اتجاه الخط العربي من اليمين إلى اليسار. | مراعاة خصائص الحرف العربي وجماله ونمط الكتابة العربية، بحيث يلتزم النص بالاتجاه من اليمين إلى اليسار، وعدم ترك فراغات بين الحروف، وأن يدعم نظام الكتابة التشكيل، وأن تظهر امتدادات الحروف -كحرف المد (مثال)- بطريقة سليمة ومتناسقة دون تشويه بصري أو انقطاع في الشكل. | نسبة الالتزام بالاتجاه | فحص واجهة المستخدم |

| المجال | المعيار العام | المعيار التفصيلي | الوصف | المؤشر | آلية القياس |
|---|---|---|---|---|---|
| | | ي. تجنُّب وجود فراغات ظاهرة بين الحروف المتصلة ببعضها. | | نسبة النصوص السليمة من الفراغات | مراجعة بصرية للنصوص |
| | | ك. دعم استخدام علامات التشكيل، أو إضافتها في حالة إمكانية إدخال نص. | | وجود خاصية التشكيل | اختبار إدخال النصوص |
| | | ل. مراعاة ظهور الحروف بشكلها الطبيعي عند استخدام الإضافات الجمالية للنص العربي كالمدود. | | نسبة الحروف الصحيحة مع الإضافات | مراجعة بصرية للنصوص بعد التنسيق |
| **ثالثًا: التداخل اللُّغوي** | ١. استخدام اللُّغة العربية ابتداءً عند بناء تطبيق متعدد اللُّغات | أ. استخدام اللُّغة العربية لغةً أساسيةً لبناء المحتوى في التطبيق. | أن تكون اللُّغة العربية حاضرةً منذ تطوير التطبيق وليست مجرد ترجمة لمحتوى كُتِب بلغة أخرى، وألا يُسبِّب المحتوى اللُّغوي تعارُضًا مع وظائف التطبيق. | نسبة المحتوى العربي الأساس | مراجعة هيكل المحتوى |
| | | ب. تجنُّب وجود تعارُض بين المحتوى اللُّغوي ووظائف التطبيق. | | عدد التعارضات المرصودة | اختبار وظائف التطبيق |

| المجال | المعيار العام | المعيار التفصيلي | الوصف | المؤشر | آلية القياس |
|---|---|---|---|---|---|
| | | ج. إمكانية التحوُّل من لغة أخرى إلى اللغة العربية على جميع صفحات التطبيق ووظائفه. | وجود المحتوى باللُّغة العربية على جميع صفحات التطبيق أو وظائفه بما في ذلك التعليمات، وإخلاء المسؤولية، وطرق التواصل. | نسبة الصفحات القابلة للتحول للعربية | قياس تجربة المستخدم في التنقل بين اللغات |
| | | د. تجنُّب الاختصارات باللُّغة العربية، مثل: (م ع س)، للدلالة على المملكة العربية السعودية. | تفادي استخدام الاختصارات نظرًا لعدم وجود توافق دائم في استخدامها بين اللغات؛ فبينما تُعد شائعة ومفهومة في اللغات الغربية، تفتقر اللغة العربية إلى نمط مماثل في الاختصار، مما قد يؤدي إلى الغموض أو ضعف الفهم لدى المستخدم العربي. | نسبة النصوص الخالية من الاختصارات | مراجعة النصوص والواجهة |
| **رابعًا: فئات المستخدمين العمرية ومستوياتهم العلمية واللُّغوية** | ١. استخدام اللُّغة المناسبة لعمر المُستخدِمين | أ. التركيز على استخدام الصورة والصوت واللُّغة البسيطة للأطفال الصغار. | يستخدم التطبيق في كل محتواه التعليمي لغةً مناسبةً للعمر المُستهدَف بالتطبيق، فيراعي توظيف الصور والأصوات بشكل مُكثَّف في التطبيقات الموجهة للأطفال، ويتدرج في مستوى التعقيد اللُّغوي كلما تقدَّم عُمر المستخدم. | نسبة الوسائط البسيطة للأطفال | مراجعة واجهة المستخدم والمحتوى |
| | | ب. التدرُّج في مستوى تعقيد اللُّغة بناءً على تقدُّم المُستخدِم في العمر. | | ملاءمة النصوص للفئات العمرية | تحليل النصوص بحسب الفئة العمرية |

| المجال | المعيار العام | المعيار التفصيلي | الوصف | المؤشر | آلية القياس |
|---|---|---|---|---|---|
| | | ج.استخدام خطوط كتابة شائعة وواضحة. | | نسبة الخطوط الواضحة | فحص بصري لواجهات التطبيق |
| | ٢.مراعاة المستوى العلمي واللُّغوي للمُستخدِمين | أ. استخدام لغة ومصطلحات مناسبة لمستوى المُستخدِمين العلمي واللُّغوي. | يراعي التطبيق المستوى العلمي واللُّغوي للمُستخدِم؛ فالتطبيقات المُوجَّهة لمَن هم في التعليم العام مختلفة عن تلك المُوجَّهة لمَن هم في التعليم العالي أو للمتخصصين والباحثين. | نسبة وضوح المصطلحات للفئة المستخدمة | استبيان ملاحظات المستخدمين |
| | | ب.بناء وظائف التطبيق وتعليماته بما يتناسب مع المستوى العلمي واللُّغوي للمستخدم. | | ملاءمة الوظائف والتعليمات | اختبار قابلية الاستخدام للمستويات العلمية واللغوية |
| خامسًا: معايير التوطين | ١.استخدام صيغة التاريخ والوقت المعيارية في بلد المُستخدِم | أ. دعم التطبيق لصيغ التاريخ والوقت المعيارية في بلد المُستخدِم. | أن يوفّر التطبيق دعمًا كافيًا لصيغ التاريخ والوقت المعتمدة في مختلف البلدان، بما يضمن توافقه مع التنسيقات المعيارية المتّبعة في بيئة المستخدم. ويشمل ذلك تنسيقات التاريخ مثل: (يوم – شهر– سنة)، أو (يوم/ شهر/ سنة)، أو (شهر/ يوم/ سنة)، وغيرها، إضافةً إلى تنسيقات الوقت المختلفة، سواء كانت بصيغة ١٢ ساعة مثل: (٠٣:١٦م)، أو بصيغة ٢٤ ساعة مثل: (١٥:١٥)، بما يتماشى مع النظام المأخوذ به في بلد المُستخدِم. | نسبة الالتزام بالتنسيق المحلي | فحص الواجهة ومطابقة الصيغة |

| المجال | المعيار العام | المعيار التفصيلي | الوصف | المؤشر | آلية القياس |
|---|---|---|---|---|---|
| | ٢. دعم المنطقة الزمنية Time Zone | أ. دعم المناطق الزمنية المختلفة حسب البلدان ولا سيما الناطقة بالعربية. | أن يدعم التطبيق المناطق الزمنية المختلفة، ويدعم التوقيت الصيفي (زيادة ساعة عن التوقيت الرسمي) في البلدان التي تُطبِّق التوقيت الصيفي، وذلك بشكل آلي؛ لضمان دقة المواعيد وتزامن الأحداث الزمنية. | توفر خيار تعديل المنطقة الزمنية | اختبار واجهة التطبيق |
| | | ب. وجود خيار يدعم التوقيت الصيفي لمنطقة المُستخدِم آليًّا في حال كانت المنطقة الزمنية مختلفة. | | توفر دعم التوقيت الصيفي آليا | اختبار واجهة التطبيق |
| | ٣. دعم التقويم الهجري والمحلي | أ. تضمين التقويم الهجري وغيره من التقاويم المحلية في بلد المستخدم. | أن يوفر التطبيق التقويم الهجري وغيره من التقاويم المحلية في بلد المُستخدِم. وكذلك استخدام أسماء الأشهر الميلادية حسب الشائع في البلد (مثل يناير مقابل جانفي، أو تشرين الثاني)، وذلك بشكل آلي حسب بلد المُستخدِم. | نسبة الشاشات الداعمة للتقويم الهجري | مراجعة إعدادات التقويم |
| | | ب. استخدام أسماء الأشهر الميلادية حسب الشائع في بلد المُستخدِم. | | نسبة الشاشات الداعمة للتقويم المحلي | اختبار عملي للإعدادات المحلية |

| المجال | المعيار العام | المعيار التفصيلي | الوصف | المؤشر | آلية القياس |
|---|---|---|---|---|---|
| | ٤.دعـم تنسيـق العـملة المحلية | أ.دعـم العـملات المختـلفة باللُّـغة العـربيـة خصوصًا الـبلـدان العـربية. | أن تدعم واجهة المُستخدِم عرض العملة المحلية حسب الشائع في بلدها، ولا سيما في البلدان العربية. وكذلك يدعم التطبيق الفواصل العشرية بحسب السائد في بلد المستخدم. | نسبة الصفحات الـتي تـعـرض الـعـمـلـة الـعـربـيـة | مـراجـعـة الـشـاشـات الـمـالـيـة |
| | | ب. تحقيق الفواصل العشرية حسب الشائع في الـبلد. | | تحقيق الفـواصل الـعـشـريـة بشـكل مناسـب | فحص الفواصل |
| | | ج. دعـم رمـوز العملات المحلية. | | دعـم رمـوز العملات المحلية | فـحـص الـرمـوز |
| **سـادسًـا: مـعـايـيـر واجـهـة المُستخدِم** | ١.تخـطـيـط واجـهـة المُستخدِم من اليمين إلى اليسار RTL | أ.اتساق الواجهة مع نظام الكتابة العربية | أن تـكـون مـحـاذاة الواجهة وجميع عناصرها متناسقةً مع الاتجاه العربي من اليمين إلى الـيـسـار: كـالـقـوائـم، والأزرار، وأشرطة الأدوات. | تحقق الاتساق | مراجعة تقنية للواجهة |
| | | ب.اتسـاق الأزرار والـقـوائـم مع الواجهة. | | تحقق الاتساق | مراجعة تقنية للواجهة |
| | ٢.تـوؤام أزرار التنـقـل في التطبيـق أو الموقع عبرواجهة المُستخدِم مع نظام الكتابة العربية | أ. وضع زر التالي عـلـى جـهـة اليسار، والسابق جهة اليمين | يجب تـوفير تجربة مُستخدِم متسقة، بحيث تكون أزرار التنقل (مثل زر التالي وزر السابق) في المواقع المتوقعة للمُستخدِم العربي. | وضع الأزرار في مكانها الصحيح | اختبار تجريبي لـوظـائـف التنقل |
| | ٣. دعم التنقل عـبرلـوحـة المـفـاتـيـح لتطبيقـات الحاسوب. | ب.دعـم التـطبيـق للتنقـل عـبرأزرار لـوحـة المفاتيح بشكل متسق بحيث يكون السهم لليسار للتالي، واليمين للسابق. | دعم تخطيطات لوحـة المفاتيح العربية، وضمان ترتيب التنقل الصحيح في سياق من اليمين لليسار (RTL). | نسبة العناصر الـتي يـتـبـع تركيزها تـرتـيـب منطقي RTL | اختبار تفاعلي بـاسـتـخـدام لوحة المفاتيح |

| المجال | المعيار العام | المعيار التفصيلي | الوصف | المؤشر | آلية القياس |
|---|---|---|---|---|---|
| | ٤. استعمال الألوان المتناسبة ثقافيًّا. | دعم الألوان التي تتناسب مع الثقافة المحلية. | توفير ألوان متناسبة مع الذوق العام والثقافة العربية، مثل: استخدام الألوان الدافئة والمحافظة التي تتناسب مع السياقات الثقافية للمُستخدِم. | نسبة الواجهات التي تستعمل ألوانًا مقبولة ثقافيًا | مراجعة تصميمية أو استطلاع رأي المستخدمين |
| **سابعًا: معايير النصوص العربية وترميزها** | ١. استخدام ترميز معياري | استخدام ترميز UTF-8 لدعم الأحرف العربية. | أن يكون الترميز المُستخدَم للغة العربية ترميزًا معياريًّا (مثل: UTF-8)، يدعم مختلف المنصَّات وأنظمة التشغيل، ويتفادى الاعتماد على أنظمة ترميز قديمة أو غير شائعة؛ الأمر الذي يضمن سلامة عرض الأحرف العربية، وعدم ظهور رموز غريبة أو مشكلات في القراءة. | نسبة الملفات أو الصفحات التي تستخدم ترميز UTF-8 | فحص تقني لترميز الملفات أو تحليل خصائص النصوص |
| | ٢. استخدام طرق إدخال تدعم العربية | أ. توفير لوحة مفاتيح عربية افتراضية. | أن يُوفَّر للتطبيق أو النظام وسائل إدخال باللُّغة العربية، مثل: لوحة مفاتيح افتراضية تحتوي على الأحرف العربية، ودعم التخطيطات القياسية الشائعة؛ مما يتيح للمُستخدِمين الكتابة باللُّغة العربية بسهولة ودون الحاجة إلى حلول خارجية أو مبدِّلات لوحة مفاتيح غير رسمية.(٢) | توفر لوحة مفاتيح عربية مدمجة | مراجعة واجهة الإدخال أو اختبار الاستخدام الفعلي |

(٢) مبدلات لوحة المفاتيح غير الرسمية: هي أدوات أو برامج خارجية يثبتها المستخدم لتمكين الكتابة بالعربية عند غياب الدعم المدمج، وتشمل برامج مثل Keyman وAutoHotkey، أو لوحات مفاتيح معدَّلة غير معتمدة من النظام. كما تشمل مواقع إلكترونية تقدم لوحات مفاتيح افتراضية أو أدوات لتحويل "عربي-فرانكو" إلى نص عربي. يُفضل تجنبها لتفادي مشكلات في الأمان والتوافق.

| المجال | المعيار العام | المعيار التفصيلي | الوصف | المؤشر | آلية القياس |
|---|---|---|---|---|---|
| | | ب. دعم التخطيطات القياسية الشائعة في لوحة المفاتيح العربية، مثل Arabic 101. | | توفر دعم التخطيطات القياسية | اختبار عملي لتخطيط لوحة المفاتيح ضمن النظام |
| | ٣. استخدام خطوط عربية مقروءة وشائعة | أ. استخدام خطوط عربية شائعة، مثل Arial. | اختيار خطوط عربية واضحة ومُوحَّدة (كخط Arial أو غيره من الخطوط القياسية) بحيث تكون حروفها متناسقةً وخاليةً من العيوب في التشكيل؛ مما يضمن سهولة التصفح والقراءة، ويقلّل من مشكلات تداخل الأحرف أو تحوُّلها إلى رموز غير مفهومة. | نسبة الخطوط العربية الشائعة | فحص الخطوط المدعومة |
| | | ب. تناسق الخطوط المُستخدَمة ووضوحها. | | نسبة الخطوط الواضحة | فحص الخطوط المدعومة |
| | ٤. استخدام ترجمة سليمة للنصوص | أ. تجنُّب استخدام الاختصارات عند الترجمة، مثل KSA لا تترجم إلى م ع س بل "المملكة العربية السعودية". | أن تُحافظ الترجمة على المعنى الأصلي وتُراعي العُرف الثقافي واللُّغوي العربي. كما يجب تفادي الاختصارات غير المألوفة في العربية، وضمان تناسق المصطلحات بين أجزاء التطبيق. ويُفضَّل كذلك الابتعاد عن الترجمة الحرفية التي قد تؤدي إلى اللبس أو تتناقض مع الأعراف اللُّغوية والسياسية المعتمدة عربيًّا. | نسبة النصوص التي تتضمن ترجمات صحيحة متوافقة مع السياق العربي | مراجعة لغوية متخصصة للترجمة |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعيار التفصيلي | الوصف | المؤشر | آلية القياس |
|---|---|---|---|---|---|
| | | ب. أن تكون الترجمة متسقةً مع المحتوى والثقافة والسياسة العربية، كالخليج العربي مثلًا. | | نسبة النصوص التي تتضمن ترجمات صحيحة متوافقة مع السياق العربي | مراجعة لغوية متخصصة للترجمة |
| | ٥. مراعاة السلامة الصرفية والنحوية | توافق النص مع الصرف العربي. | التحقُّق من سلامة النصوص المُضمَّنة لغويًّا؛ خصوصًا صيغ الجمع والمثنى، وقواعد التذكير والتأنيث، وأزمنة الأفعال، مثل: "قطعتين / قطعتان" بدلًا من "٢ قطعة". | نسبة النصوص الخالية من الأخطاء الصرفية والنحوية | تدقيق لغوي يدوي أو باستخدام أدوات معالجة لغوية |
| | ٦. استعمال الفُصحى في النصوص المُستعملة. | كتابة جميع النصوص في التطبيق باللُّغة العربية الفُصحى المعاصرة MSA. | يجب أن تُكتَب الواجهات والنصوص الأساسية بلغة عربية فُصحى واضحة؛ لتُيسِّر على مختلف شرائح المستخدمين فهمها. كما يجب تجنُّب اللهجات المحلية أو العبارات العاميَّة ما لم يكن لذلك هدف واضح يناسب سياق التطبيق وفئته المستهدفة. | نسبة النصوص المكتوبة بلغة عربية فصحى واضحة | تحليل لغوي نوعي أو مراجعة بشرية للنصوص |
| | ٧. استعمال خصائص العربية في البحث | استعمال المرادفات والتصريفات في بحث التطبيق. | أن يدعم نظام البحث في التطبيق أو الموقع معالجةً لغوية تتيح فَهْم المرادفات، والتصريفات المختلفة مما يُعزِّز إمكانية الوصول إلى المحتوى الملائم لاحتياجات المُستخدِم. | نسبة استجابات البحث التي تستفيد من المرادفات أو التصريفات | اختبار وظيفة البحث باستخدام كلمات متنوعة لغويًا |

| المجال | المعيار العام | المعيار التفصيلي | الوصف | المؤشر | آلية القياس |
|---|---|---|---|---|---|
| **ثامنًا: معايير استعمال النماذج اللُّغوية** | ١. استعمال معالجات نصية مُعتمدة | أ. استعمال المعالجات النصية الشائعة. | استعمال معالجات نصية مُعتمدة موثوقة، مثل: فراسة Farasah، وتقسيم Takseem. استخدام أدوات معالجة نصية (تقطيع ومحللات صرفية، وغيرها) مُختبرة في بيئات بحثية أو صناعية، والتأكد من استخدام نُسخ مُحدَّثة تعالج أخطاء الإصدارات القديمة. كما يُفضّل التحقُّق دوريًّا من مستوى دقة الأدوات، ولا سيّما عند التعامل مع لهجات متنوعة أو أنواع نصية مختلفة (أدبية، علمية...). | نسبة استخدام أدوات معالجة نصية موثوقة | مراجعة قائمة الأدوات المستخدمة |
| | | ب. استخدام إصدارات مُحدَّثة من الأدوات المعتمدة للغة العربية. | | نسبة الأدوات المستخدمة بإصداراتها المحدثة | فحص تقني لإصدارات الأدوات |
| | ٢. معالجة النصوص قبلَ التدريب | أ. تنقية النصوص من الأخطاء. | في بعض التطبيقات، قد يُفضّل إزالة علامات التشكيل لتفادي تشتُّت النموذج أو للتركيز على الأبجدية الأساسية؛ في حين أنه في تطبيقات أخرى (مثل تدريب نموذج لضبط الإعراب) يُنصح بالإبقاء على التشكيل. ويجب تحديد الغرض مُسبقًا، وضبط إجراءات التنقية قبل مرحلة التدريب. | التحقق من سلامة النصوص | مراجعة النصوص |

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

| المجال | المعيار العام | المعيار التفصيلي | الوصف | المؤشر | آلية القياس |
|---|---|---|---|---|---|
| | | ب. إزالة علامات التشكيل عند الرغبة في تدريب نموذج يتجاهل الحركات. | | التحقق من إزالة علامات الترقيم غير اللازمة | مراجعة النصوص |
| | ٣. استعمال المُقسّمات المعتمدة | أ. استخدام تقنيات فصل النص الشائعة. | ينبغي أن يدعم المُقسّم المعتمد (مثل BPE أو SentencePiece) الخصائص الفريدة للعربية (الاتصال بين الحروف، اللواحق، السوابق...). قد يستلزم الأمر تدريب المُقسّم على بيانات عربية واسعة النطاق؛ لضمان دقة أكبر في تجزئة الكلمات وتعريف الوحدات الصرفية. | مدى دقة تجزئة الكلمات باستخدام المقسّم | اختبار على نصوص عربية متنوعة |
| | | ب. التحقُّق من كون المُقسّم مُهيَّأً للغة العربية. | | نسبة الخصائص العربية المدعومة في المقسّم | تحليل دعم المقسم للخصائص العربية |
| | ٤. استعمال النماذج الداعمة للعربية | أ. اعتماد نماذج لغوية عربية مُتخصِّصة بدلًا من استخدام نماذج عالمية لا تتضمَّن تمثيلات كافية للعربية. | نظرًا لقلّة عدد النماذج العربية عالية الجودة مقارنةً باللُّغات العالمية؛ ينبغي الحرص على اختيار النماذج التي تناسب الخصائص اللُّغوية للعربية، وتحافظ على مستوى معقول من الدقة. كذلك يُوصَى بتحديث النموذج عند صدور إصدارات جديدة منه تدعم مزايا إضافية أو تُحسِّن أداءَه مع اللهجات المختلفة. | نسبة النماذج العربية المتخصصة المستخدمة | تحليل نوعية النماذج المستخدمة |

| المجال | المعيار العام | المعيار التفصيلي | الوصف | المؤشر | آلية القياس |
|---|---|---|---|---|---|
| | | ب. متابعة تحديثات النماذج العربية وتحسيناتها. | | نسبة النماذج المستخدمة بإصدارها الأحدث أو المدعّم بتحسينات | مراجعة سجل الإصدارات وتاريخ الاستخدام |
| **تاسعًا: معايير الأداء والصيانة** | ١. التحقُّق من جودة أداء التطبيق وسرعة استجابته للأوامر. | أ. التأكُّد من توافق التطبيق مع المحتوى العربي. | مراجعة بنية التطبيق، والتأكُّد من توافقه مع المحتوى العربي من دون تباطؤ ملحوظ. على سبيل المثال، قد يستلزم الأمر استخدام آليات تتعامل مع تعدُّد الصيغ الحرفية[٣] والاتجاه النصي (RTL)، مع تفادي حلول مؤقتة تستهلك وقتًا إضافيًا أو تؤدي إلى أعطال. | نسبة الأنظمة التي تعمل بكفاءة مع المحتوى العربي | اختبار جودة الأداء عند التعامل مع بيانات عربية |
| | | ب. ضبط جودة أداء التطبيق وسرعة استجابته للأوامر. | | نسبة العمليات التي يتم تنفيذها في الزمن المتوقع | اختبار الأداء عبر محاكاة الاستخدام الحقيقي |
| | | ج. ضبط خوارزميات التحميل والمعالجة عند التعامل مع النصوص العربية. | | نسبة الخوارزميات المحسنة للتعامل مع العربية | تحليل الرمز البرمجي أو مراجعة تصميم المعالجة |

(٣) "الصيغ الحرفية" تشير إلى اختلاف شكل الحرف الواحد اعتمادًا على موقعه في الكلمة، وهي من الخصائص الفريدة للغة. فالحرف يتغيّر شكله إذا جاء في أول الكلمة، أو وسطها، أو آخرها، أو منفصلًا. على سبيل المثال، حرف "عين (ع)" يظهر بصيغ مختلفة: عـ (أول الكلمة، مثل: عَلَم)، ـعـ (وسط الكلمة، مثل: مَعَلَم)، ـع (آخر الكلمة، مثل: يمنع)، وع (منفصلًا). دعم هذه الصيغ ضروري لعرض النص العربي بشكل صحيح.

| المجال | المعيار العام | المعيار التفصيلي | الوصف | المؤشر | آلية القياس |
|---|---|---|---|---|---|
| | ٢. مراعاة دعم اللُّغة العربية في أساس برمجة التطبيق | أ. دعم اللُّغة العربية في مراحل البرمجة الأولى. | وضع خطة واضحة لدعم العربية منذ البداية، بدلًا من إضافتها لاحقًا كترجمة أو تعديل ثانوي؛ تجنُّبًا للتعارض مع مكوِّنات التطبيق الأخرى أو تعطيل أداء بعض الوحدات. | نسبة المكونات التي تدعم العربية منذ التأسيس | مراجعة توثيق مراحل التطوير |
| | | ب. اعتماد بنية قادرة على التوسُّع لدعم نُسخ / لهجات عربية مستقبلًا. | | مدى جاهزية البنية لتوسعة دعم العربية | تحليل معماري للنظام أو المنتج |
| | ٣. توفير الدعم الفني بالعربية | أ. تقديم قنوات دعم فني باللُّغة العربية (هاتف، بريد إلكتروني، منصة تذاكر[٤]...). | إتاحة الدعم الفني بالعربية للمُستخدِمين، وتضمين صفحات الأسئلة الشائعة (FAQ) بالعربية، أو إنشاء خطوط هاتفية مُتخصِّصة لحل مشكلات المُستخدِم؛ مما يُعزز ثقة العميل في المنتج وسهولة تعامله معه. | نسبة قنوات الدعم المتوفرة بالعربية | اختبار فعلي لقنوات الدعم أو مراجعة وثائقها |
| | | ب. توفير معلومات إرشادية ووثائق مساعدة باللغة العربية. | | نسبة الوثائق المساندة المتوفرة بالعربية | مراجعة محتوى قسم المساعدة والتوثيق |

(٤) منصة التذاكر هي نظام إلكتروني لإدارة طلبات الدعم الفني، حيث يتم إنشاء "تذكرة" لكل استفسار أو مشكلة يقدّمها المستخدم، ليتم متابعتها وحلّها من قبل فريق الدعم. تتيح هذه المنصات – مثل Zendesk وFreshdesk – تنظيم الطلبات وتوثيقها وتحسين جودة الخدمة بشكل فعّال.

| المجال | المعيار العام | المعيار التفصيلي | الوصف | المؤشر | آلية القياس |
|---|---|---|---|---|---|
| **عاشرًا: معايير التوافقية بين المنصات المختلفة والتحقُّق من المحتوى** | ١. اختبار واجهة المُستخدِم عبر أنظمة تشغيل مختلفة | أ. اختبار الواجهة على أنظمة تشغيل مختلفة وشائعة، بإصداراتها المختلفة. | اختبار التطبيق عبر أنظمة التشغيل الشائعة مثل: Windows وMacOS، والأجهزة الذكية مثل: أنظمة Android وiOS وغيرها، وأن تكون مدعومةً على جميع الأجهزة وأنظمة التشغيل الشائعة. | نسبة الأنظمة المدعومة من التطبيق | اختبار تجريبي على الأجهزة وأنظمة التشغيل |
| | | ب. اختبار الواجهة على أجهزة مختلفة. | | نسبة الأجهزة التي تظهر الواجهة بشكل صحيح | تجربة فعلية لتوافق الواجهة |
| | | ج. دعم المنصات والأجهزة الشائعة من دون مشكلات. | | نسبة المنصات المدعومة بالكامل | تحليل النتائج عبر تقارير الأداء |
| | ٢. اختبار واجهة المُستخدِم عبر المتصفحات المختلفة | أ. اختبار الواجهة على متصفحات مختلفة وشائعة بإصداراتها المختلفة. | التحقُّق من توافق العرض والأداء في أشهر المتصفحات، والتأكد من صحة اتجاه النص العربي (RTL) وعدم وجود مشكلات في الخطوط أو التشكيل. ويُستحسن أيضًا فحص الإصدارات الحديثة والسابقة نسبيًّا من كل متصفح؛ لتفادي الأخطاء المرتبطة بدعم اللُّغة العربية. | نسبة المتصفحات المدعومة بدون خلل في النص العربي | اختبار الأداء عبر المتصفحات |
| | | ب. اختبار الواجهة على أحجام شاشات مختلفة. | | نسبة الشاشات التي تتوافق مع تصميم الواجهة | مراجعة توافق التصميم المتجاوب |

## المراجع :


- الثبيـتي، عبـد المحسـن. (٢٠٢١م). دليـل توسـيم البيانـات اللُّغويـة. منشـورات مدينـة الملـك عبـد العزيـز للعلـوم والتقنيـة.
- العصيـمي، صالـح فهـد؛ وآخـرون. (Ed.). (٢٠١٥م). المدونـات اللُّغويـة العربيـة: بناؤهـا وطرائـق الإفـادة منهـا.
- الفيفـي، عبـد الله يحـيى (٢٠٢٣م). وسـم المدونـات اللُّغويـة: المفهـوم والمجـالات. سلسـلة العلـوم الإنسـانية والاجتماعيـة، ١(٣٨)، ص ٢٧٥-٣١٦.

- Atkins, S., Clear, J., & Ostler, N. (1992). Corpus design criteria. *Literary and Linguistic Computing, 7*(1), 1–16. https://doi.org/10.1093/llc/7.1.1
- Belinkov, Y., Habash, N., Kilgarriff, A., Roth, R., & Suchomel, V. (2013). arTenTen: a new, vast corpus for Arabic. *In Proceedings of WACL* (p. 20).
- Garside, R., & Leech, G. N. (Eds.). (1997). *Corpus annotation: Linguistic information from computer text corpora*. Longman.
- Ide, N. (1998). Corpus Encoding Standard: SGML Guidelines for Encoding Linguistic Corpora. *In LREC* (pp. 463–470).
- Ide, N., & Pustejovsky, J. (Eds.). (2017). *Handbook of linguistic annotation*. Springer Netherlands.
- Ide, N., & Suderman, K. (2014). The linguistic annotation framework: a standard for annotation interchange and merging. *Language Resources and Evaluation, 48,* 395–418.
- Jarrar, M., Khalilia, M., & Ghanem, S. (2022). Wojood: Nested Arabic named entity corpus and recognition using BERT. *In Proceedings of the Thirteenth Language Resources and Evaluation Conference* (pp. 3626–3636).

هذه الطبعة إهـداء من المجمـع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً


- Lai, V. D., Ngo, N. T., Veyseh, A. P. B., Man, H., Dernoncourt, F., Bui, T., & Nguyen, T. H. (2023). *ChatGPT beyond English: Towards a comprehensive evaluation of large language models in multilingual learning*. arXiv preprint arXiv:2304.05613.
- Maamouri, M., Bies, A., Buckwalter, T., & Mekki, W. (2004). The Penn Arabic Treebank: Building a large-scale annotated Arabic corpus. *In NEMLAR conference on Arabic language resources and tools* (Vol. 27, pp. 466–467).
- Marcus, M., Santorini, B., & Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics, 19*(2).
- Pustejovsky, J., & Stubbs, A. (2012). *Natural language annotation for machine learning.* O›Reilly Media, Incorporated.

هذه الطبعة **إهـداء من المجمـع**
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

# شكر وتقدير

يشـكر مجمـع الملـك سـلمان العالـمي للغـة العربيـة فريـق العمـل الذيـن عملـوا علـى التخطيـط لبنـاء هـذه المعايـير، وتأليفهـا وتحكيمهـا ومراجعتهـا. **وهـم:**

**فريق الإعداد العلمي:**

د. أيمن بن أحمد المنصوري.

د. صخر بن بدر الخريف.

د. أحمد بن علي الأخشمي.

**فريق التخطيط والتحكيم والمراجعة:**

أ.د. محمود بن عبد الله المحمود.

أ.د. خالد بن سليمان القوسي.

د. خالد بن عبد الرحمن الميمان.

د. سارة بنت عمر الحمود.

د. نجلاء بنت عبد الرحمن الزامل.

أ. نهال بنت علي القحطاني.

أ. مياسم بنت قنان المطيري.

أ. أسيل بنت عبد العزيز السيف.

مجمع الملك سلمان العالمي للغة العربية
King Salman Global Academy for Arabic Language



هذه الطبعة
إهداء من المجمع
ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

