مكنز «محيط»: منهجية بناء مكنز دلالي موثّق للتراث الإسلامي العربي والفارسي

شادي علي

صادقون للتقنية — Sadiqoon Technologies Ltd، لندن

مسجّلة في إنجلترا وويلز برقم 17120720 · www.sadiqoon.uk

info@sadiqoon.uk

تشرين الثاني / نوفمبر 2026

ورقة منهجية — الإصدار 1.0

«مكنز محيط» خدمة من خدمات صادقون للتقنية · moheet-ablibrary.net

الملخص

تعرض هذه الورقة منهجية بناء «مكنز محيط» (Moheet Corpus): مكنز دلالي موثّق يستوعب مدونة «مكتبة أهل البيت عليهم السلام» بلسانَيها العربي والفارشي، ويقوم على وحدة بناء واحدة هي «المقطع الموثّق»: قطعة نصية محدودة الطول تحمل معها بطاقة نسبها كاملة (الكتاب، المؤلف وسنة وفاته، الجزء، الصفحة، سلسلة العناوين، وطبيعة النص). تمرّ الكتب بهضم بنيوي يحفظ هوامش المحققين في مواضعها، ثم تُقطَّع على حدود المعنى، وتُمثَّل بمتجهات متعددة اللغات بقياس 1536 بُعدًا إلى جانب تمثيل لفظي متفرق، ويُسترجع منها بدمج رتبي ثم إعادة ترتيب. وفوق الفضاء المتجهي يُبنى غراف معرفي لا تُقيِّد فيه علاقة بين مفهومين أو علمين أو كتابين إلا ومعها دليلها الحرفي وموضعه بالصفحة.

بلغ المكنز في تشرين الثاني 2026 نحو 13.1 مليون مقطع موثّق من قرابة سبعة آلاف مصنّف، من أصل ثمانية عشر ألفًا تضمّها المكتبة. وتُقاس في الورقة ثلاث خصائص: دقة العزو إلى الصفحة المطبوعة (99.99% على عينة مقيسة من عشرة كتب)، ومنع تكرار الكتاب الواحد عبر بصمة الملف ومعرّفه في المكتبة، وكلفة الهضم (نحو ثلاثة سنتات للكتاب في التضمين، أي دولار واحد لكل عشرة آلاف صفحة). وتختم الورقة بحدود المنهج وبالتزامٍ بإتاحة مفتوحة: الغراف الببليوغرافي للمكتبة تحت رخصة مفتوحة، ومعيار علني لاستخراج الكيانات والعلاقات من النثر العلمي العربي الكلاسيكي.

الكلمات المفتاحية: التراث الإسلامي الرقمي، الاسترجاع الهجين، الغراف المعرفي، التوثيق الببليوغرافي، بروتوكول سياق النماذج (MCP)، العربية والفارسية.

1. المقدمة: المشكلة وسؤال الورقة

مرّت رقمنة التراث الإسلامي بجيلين: جيل الحفظ بالتصوير، ثم جيل النص المفهرس والبحث اللفظي الذي تمثّله المكتبات الرقمية الكبرى اليوم. وفي الجيل الثاني يجد الباحث النص إن عرف لفظه، ولا يجده إن عرف معناه وحده؛ ولا يرى العلاقات التي تربط مسألةً في كتاب فقهي بأصلها في كتاب حديثي وبنقدها في شرح متأخر، إلا بجهد بشري يستغرق العمر. وقد دخلت النماذج اللغوية الكبيرة هذا الحقل بوعد الفهم وبخطر الاختلاق: نصوص وعزوات لا وجود لها تُقال بثقة لغوية كاملة \[1\]، وهو خطر لا يُحتمل في علوم يُتعبَّد فيها بالنص.

سؤال الورقة منهجي محض: كيف تُبنى بنية معرفية دلالية فوق مدونة موسوعية متعددة الفنون واللغات، بحيث يستحيل بنيويًا أن يصل إلى الباحث نص بلا عزو، أو علاقة بلا دليل؟ الجواب الذي تعرضه الورقة ليس تحسينًا لنموذج، بل قيد تصميمي يسري في الطبقات كلها: التوثيق جزء من مادة البناء لا طبقة تُضاف عند العرض.

وتقع الورقة بين عملين سابقين للمؤلف: الورقة التعريفية «مكنز محيط: نقلة في هيكلية المعارف الإسلامية الشيعية» (أيلول 2026) \[2\] التي تناولت موقع المشروع على طريق الاجتهاد وأثره في الدرس، ومعيار الاستخراج المفتوح الذي تمهّد له هذه الورقة. فهي تقتصر على المنهج: ما الذي بُني، وكيف، وبأي قيود، وما الذي قيسَ.

وتستأنس بمنهج مشاريع رقمنة التراث العربي الأكاديمية، ولا سيما OpenITI ومشروع KITAB في توحيد أسماء المؤلفين وتمييز الطبعات \[3\]، وبمشاريع غرافات الحديث المحكّمة \[4\]\[5\]، وتختلف عنها في أمرين: اتساع المدونة إلى أربعة عشر فنًا بلسانين، وقيد الدليل الحرفي في الغراف.

2. المدونة ووحدة البناء

المدونة المستوعبة هي مدونة برنامج «مكتبة أهل البيت عليهم السلام» بنسخها الرقمية المحققة: نحو ثمانية عشر ألف مصنّف عربي وفارشي، بحجم يناهز أربعة وعشرين جيجابايت من النصوص، موزعة على أربعة عشر فنًا: الحديث، والفقه، وأصول الفقه، والتفسير، والعقائد والكلام، والرجال، والأخلاق، والتاريخ، والدعاء، والمعاجم، والشروح العلمية، والفلسفة والمنطق، واللغة والنحو والبلاغة، والأدب. والاستيعاب كامل لا انتقائي، لأن قيمة البنية الدلالية تأتي من العلاقات عبر الفنون لا من داخل الفن الواحد؛ وللثنائية اللغوية دلالة منهجية، إذ كُتب قسم وازن من تقريرات الدروس والشروح في القرون الأخيرة بالفارسية وظل شبه مقطوع عن الباحث العربي، والتمثيل الدلالي يضع التراثين في فضاء واحد يُسأل باللغتين.

تدخل الكتب من ملفات المكتبة المصدّرة بوسومها البنيوية الأصلية: فواصل الصفحات المطبوعة، وكتل الهوامش، وفهارس الموضوعات، وبطاقة الكتاب (المؤلف وسنة وفاته، المحقق، الطبعة، الناشر، الجزء). وتُقرأ البطاقة آليًا من الملف نفسه فلا يُدخل التوثيق يدويًا فيُخطأ فيه. وقد أثبت جرد شامل لأربعة عشر ألفًا وستمائة وتسعين ملفًا أن علامة الصفحة في تصدير المكتبة تطابق المطبوع في الملفات كلها، وهو ما يجعل العزو إلى الصفحة المطبوعة ممكنًا من المصدر لا بالتقدير.

أما وحدة البناء فليست الكتاب ولا الصفحة، بل «المقطع الموثّق»: قطعة نصية محدودة الطول (وسيطها نحو 350 وحدة لغوية) تحمل معها في بياناتها الحقول التالية، ولا تنفك عنها في أي مسار لاحق:

الحقل	ما يحمله	مصدره
الكتاب والمؤلف وسنة وفاته	النسبة الببليوغرافية	بطاقة الكتاب في الملف
الجزء والصفحة (بداية ونهاية) ومواضع فواصل الصفحات داخل المقطع	العزو إلى المطبوع	وسوم الصفحات
سلسلة العناوين (الكتاب ← الباب ← الفصل)	الموضع من بنية الكتاب	الفهرس والعناوين المحقونة
رقم الحديث ورقم الباب حيث وُجدا	الإحالة الداخلية المعتادة في كتب الحديث	بدايات الأسطر بنيويًا
طبيعة النص	متن، متن بهوامش، شرح، تعليقة محقق	وسوم الهوامش والشروح
معرّف الملف في المكتبة وبصمة SHA-256	هوية الطبعة ودورة حياتها	الملف نفسه

بهذا التصميم يستحيل بنيويًا أن يظهر في المنظومة نص سائب بلا عزو؛ فالتوثيق خاصية في المقطع منذ ولادته، وكل ما يُبنى فوقه — متجه، أو علاقة في الغراف، أو جواب أداة — يرث العزو نفسه.

3. الهضم البنيوي: من ملف المكتبة إلى مقاطع تحمل توثيقها

يجري الهضم في أربع مراحل متتابعة، والقيد الحاكم فيها واحد لا تُجرى عملية تُسقط معلومة بنيوية كانت في الملف الأصلي.

1. **التحليل البنيوي للملف.** تُستخرج بطاقة الكتاب، وتُبنى الهيراركية كاملة (الكتاب ← الأبواب ← الفصول)، وتُحقن عناوين الفهرس الغائبة عن المتن في مواضعها، وتُزال وسوم التصدير التي لا تحمل معنى للقارئ (وسوم الارتباط والملحقات والشعر الخام). وتُحدد هوية الطبعة بمعرّف الملف في المكتبة لا بالعنوان والجزء وحدهما، لأن العناوين تتشابه في التراث (أمالي المفيد وأمالي الصدوق وأمالي الطوشي مثالًا).
2. **دمج الهوامش في مواضعها.** كل هامش مرقّم يُدمج في المتن عند موضع إحالته بالضبط موسومًا ‹هامش: ...›، فتبقى ترجمة الراوي ملتصقة بموضعها من السند أينما وقع التقطيع؛ وتُميّز تعليقات المحقق النجمية ‹تعليقة محقق: ...› والشروح ‹شرح: ...›. والغاية أن يصل إلى الباحث نسيج الكتاب المحقق كما طُبع، لا متنًا مجرّدًا من عمل محققيه.
3. **التقطيع الدلالي.** يُقسم النص مقاطع على حدود المعنى (اكتمال الفقرة والجملة) مع تراكب محسوب بين المقاطع المتجاورة كي لا تنقطع الفكرة عند الحد. ولأن فاصل الصفحة قد يقع داخل المقطع، يحمل المقطع صفحة بدايته ونهايته ومواضع الفواصل داخله، فتُعزى العبارة المسترجعة إلى صفحتها هي لا إلى صفحة المقطع عمومًا («ص 62–63» عند الامتداد). ولا يُسقط التقطيع الأسطر القصيرة، لأن فيها أرقام الأحاديث وعناوين الأبواب وعلامات الصفحات.
4. **التحقق والتخزين.** تُحسب لكل ملف بصمة SHA-256 تحكم دورة حياته: النسخة المكررة تُرفض، والنسخة الأحدث من الكتاب تحل محل القديمة حلولًا ذريًا، والنصية المحققة تعلو على المصورة، ولا يُحذف أصل من الأرشيف بل يُوسم ويُحفظ أثرًا للتدقيق. ولا يُهضم الكتاب الواحد مرتين ولو اختلفت طبعته، وإن وُجد أصل عربي للكتاب لم يُرفع نظيره الفارشي. وتُحفظ صور صفحات الكتب المصورة في مخزن كائنات، ولا يخزن الفهرس المتجهي إلا المتجه والمرجع.

تعمل المراحل الأربع في مسارات متوازية تحجز الكتب حجزًا ذريًا من طابور واحد، ويُسجّل لكل كتاب في دفتر الكتب ما جرى له: قُبل أو تُخطي أو رُفض، ولماذا، وبأي كلفة.

4. التمثيل المتجهي والاسترجاع الهجين وإعادة الترتيب

يُمثِّل كل مقطع موثّق تمثيلين متكاملين يُخزنان معًا في قاعدة Qdrant المتجهية \[6\]:

• **التمثيل الكثيف (المعنى):** متجه بـ 1536 بُعدًا ينتجه نموذج تضمين تجاري متعدد اللغات والوسائط (Cohere Embed v4) بنمط «وثيقة بحث» للمقاطع ونمط «استعلام» للأسئلة، بمقياس تشابه جيبي. المتقارب معنى يتجاور في هذا الفضاء وإن لم يشترك في لفظ، وهو ما أثبتته أدبيات الاسترجاع الكثيف \[7\]\[8\]، ويجعل السؤال المصوغ بلغة اليوم يصل إلى نص كُتب قبل ألف سنة، والسؤال العربي يصل إلى النص الفارشي.
• **التمثيل المتفرق (اللفظ):** متجه BM25 متفرق يمسك المصطلح الحوزوي الدقيق والعبارة بعينها، وهو المطلوب في التخريج ومقابلة الروايات.

في الاستعلام يجري البحثان معًا، ثم تُدمج القائمتان بخوارزمية دمج الرتب المتبادل (RRF) التي تعتمد موقع النتيجة في كل قائمة لا درجتها، فتتفادى تفاوت المقاييس بين المحركين \[9\]، ثم يمرّ المرشحون على نموذج إعادة ترتيب يقارن كل مرشح بالسؤال مقارنة مباشرة أدق من المقارنة الأولى. وخلف السلسلة تعمل خوارزمية الجوار التقريبي HNSW \[10\] التي تجعل البحث في 13 مليون مقطع مسألة أجزاء من الثانية.

ثلاثة قرارات منهجية تستحق التصريح:

1. **ثبات نموذج التضمين طوال الهضم.** المتجهات الناتجة من نماذج مختلفة لا تُمزج في فضاء واحد، فأي تبديل للنموذج أو للأبعاد يعني إعادة تضمين المدونة كلها. لذلك يُثبّت النموذج من بداية المشروع إلى اكتماله، وتُعامل المتجهات في سياسة النسخ الاحتياطي معاملة الأصول لا البيانات المشتقة الرخيصة.
2. **الترشيح بالبيانات الببليوغرافية جزء من الاستعلام.** لأن كل مقطع يحمل سنة وفاة مؤلفه وفنّه وكتابه، يستطيع الباحث حصر البحث في من توفي قبل سنة هجرية معينة أو بعدها، أو في فن بعينه، وهو باب لدراسات تطور المفاهيم عبر العصور كانت تتطلب من الباحث أن يبني بنفسه مكتبة مرتبة على الطبقات.
3. **النص المسترجع هو النص المطبوع.** لا تعيد طبقة العرض صياغة المقطع ولا تلخّصه، بل تعرض النص الوفي بهوامشه ورقم حديثه وبابه، مع العزو إلى الصفحة المطبوعة وإلى ترتيب النسخة المصورة حيث يختلفان.

5. الغراف المعرفي المقيّد بالدليل الحرفي

فوق الفضاء المتجهي يقوم غراف معرفي في قاعدة Neo4j، عقده المفاهيم والأعلام والكتب والرواة والأحداث، وحوافه علاقات موسومة بأنواعها من مخطط علاقات مغلق وموثّق (يستند إلى، يعارض، نقد، يتفرع عن، شرح، روى عن، ونظائرها)، مع طبقات مخطط خاصة بكل فن (فالعلاقات في كتب الرجال غيرها في كتب الفلسفة). وقد صارت الغرافات المعرفية البنية المعيارية لتمثيل المعارف المترابطة \[11\]، ودخلت الدراسات الإسلامية من باب غرافات الحديث المحدودة بمدونة بعينها \[4\]\[5\]؛ ويمد «محيط» المنطق نفسه على مدونة موسوعية بلسانين.

القيد التصميمي الحاكم لا تُقيِّد في الغراف علاقة إلا ومعها دليلها الحرفي من النص وموضعه بالكتاب والصفحة. فالغراف لا «يرى» آراء، بل يحمل نصوصًا متعالقة، وحين يُسأل عن علاقة المفهوم أ بالمفهوم ب أجاب بعبارات المصنفين أنفسهم معزوة.

منهج الاستخراج. تُستخرج الكيانات والعلاقات من كل مقطع بنموذج لغوي مفتوح الأوزان يعمل في دفعات أسبوعية على واجهة استدلال سحابية، وفق تلقين يقصره على العلاقات المصرّح بها في المقطع ويلزمه بإعادة العبارة الدالة بحرفها ثم يُتحقق من العبارة بمطابقتها على المقطع المصدر، فما لم يُطابق لا يدخل الغراف. وتُوحّد الكيانات المترادفة (العلم الواحد بأسمائه وكناه وألقابه) بخطوة دمج لاحقة تحتفظ بالأسماء البديلة كلها. ويُراقب نمو الغراف بمعيار واحد لا يُخدع: عدد الإشارات الجديدة الموثّقة في القاعدة، لا عدد المقاطع المعالجة.

قياس الجودة قبل أي تغيير. للمشروع مجموعة ذهبية من 150 مقطعًا موزعة على الفنون، موسومة بكياناتها وعلاقاتها، وأداة مقارنة A/B تقيس لأي نموذج أو تلقين جديد ثلاثة أشياء: استرجاع الكيانات،

واسترجاع العلاقات، ووفاء الدليل الحرفي للنص لا يُبدّل نموذج الاستخراج ولا تلقينه في الإنتاج إلا بعد أن يساوي أو يفوق القائم على هذه المجموعة. وتُنشر المجموعة والأداة معيارًا مفتوحًا (القسم 8).

العناقيد الموضوعية. فوق الغراف تعمل خوارزمية كشف مجتمعات من عائلة Leiden \[12]\ تبلور جزر المفاهيم شديدة الترابط، ويُولّد لكل عنقود ملخص موضوعي، على نمط الاسترجاع الغرافي المعزز \[13]\، للأسئلة الكلية التي لا يكفي فيها استرجاع مقاطع متفرقة.

حالة الغراف في تشرين الثاني 2026: الاستخراج مكتمل لنحو 5% من المقاطع ويتقدم بوتيرة تحكمها كلفة الاستدلال، وهو موضع العمل القادم (القسم 8).

6. الإتاحة: الموصل البحثي والأدوات وعمارة الصدق

لا يُقدّم المكنز موقعًا مغلقًا، بل موصلًا برمجيًا على «بروتوكول سياق النماذج» (MCP) \[14]\، المعيار المفتوح لوصل النماذج اللغوية بمصادر المعرفة، بحيث يستطيع أي مساعد بحثي أو وكيل اصطناعي مخاطبة أدواته والتركيب عليها في مسار عمل واحد. يعرض الموصل إحدى عشرة أداة قراءة تغطي دورة العمل العلمي:

الأداة	وظيفتها	الطبقة التي تعتمد عليها
البحث في المدونة	بحث دلالي عام بمعاملات ضبط (الفن، المؤلف، سنة الوفاة، الكتاب)	الاسترجاع الهجين
الاستكشاف الغرافي	العلاقات الموثّقة بين المفاهيم والأعلام والكتب مع الدليل الحرفي	الغراف
العناقيد الموضوعية	مسح بانورامي للمحاور الكبرى المتصلة بمجال	العناقيد
بطاقة الكتاب	البيانات الببليوغرافية والطبعة والمحقق	سجل الكتب
تخريج الروايات	مواضع الرواية مرتبة بأسبقية وفاة المؤلف عند الطلب، داخل نطاق المطابقة القوية حصرًا	الاسترجاع الهجين
التحقق من النصوص الحديثية	حكم صريح: موجود / ورد ما يشبهه / غير موجود بهذا اللفظ، مع درجة تطابق وحارس لاختلاف الأعداد	الاسترجاع الهجين
المفسر القرآني والشرح العلمي	استعلامات موجّهة بفنّي التفسير والشروح	الاسترجاع الهجين
القراءة المتسلسلة وعرض الصفحة	قراءة نطاق صفحات (ثلاث افتراضًا) وصورة الصفحة الأصلية برابط موقّع مؤقت	المخزن والرؤية

ثلاث خصائص تعكس أصول الصنعة لا مجرد براعة برمجية: التقييد الزمني بوفيات المؤلفين، وتفضيل أقدم المصادر في التخريج مع حارس يمنع أن يتقدم مصدر ضعيف المطابقة لمجرد قدمه، والحكم الصريح في التحقق بدل عرض المتشابهات.

عمارة الصدق. جواب المكنز على ظاهرة الاختلاق \[1\] ليس تحسينًا للنموذج بل تقييد معماري له على منهج الاسترجاع المعزز \[15\]، ثم يتجاوزه بجعل مخرجات كل أداة محكومة بعتبات مطابقة صريحة: تعود المنظومة بجواب «لا توجد نتائج ذات صلة» أو «لا توجد علاقات موثّقة» متى قصرت المطابقة عن العتبة، وتعد الجواب الفارغ نتيجة صادقة لا عطلًا. ووظيفة الآلة في هذا كله التنبيه لا الإفتاء: هي مصنع قرائن يعرض النصوص بعزوها، والترجيح للباحث.

ويسري على الموصل ضبط تشغيلي: وصول موثّق بحسابات، وحصص استخدام تُقاس بالنداء ونوعه، وسجل نفاذ، ولا تصدير للمدونة بالجملة.

7. التقييم المقيس

الأرقام التالية مقيسة من المنظومة الحية لا مقدّرة، وتاريخ القياس مذكور لكل منها.

الخاصية	القيمة المقيسة	العينة والتاريخ
حجم المكنز	13,127,394 مقطعًا موثّقًا، متجه كثيف 1536 بُعدًا + متجه متفرق لكل مقطع	المجموعة الحية، 1 تشرين الثاني 2026
المصنفات المهضومة	نحو 7,000 من نحو 18,000	دفتر الكتب، تشرين الثاني 2026
دقة العزو إلى الصفحة المطبوعة	99.99% من المقاطع تُعزى إلى صفحتها الصحيحة	10 كتب مضاهاة بموقع المكتبة، أيلول 2026
مطابقة علامة الصفحة للمطبوع في ملفات المكتبة	100% (14,690 من 14,690 ملفًا)	جرد كامل، أيلول 2026
استرجاع أرقام الأحاديث والأبواب بنيويًا	98.9%–100%	كتب حديثية مختارة، أيلول 2026
طول المقطع	وسيط 356 وحدة لغوية	3,676 كتابًا، أيلول 2026
كلفة التضمين للكتاب الواحد	وسيط 0.031 دولار (متوسط 0.033، المئين 90: 0.060) ≈ دولار لكل عشرة آلاف صفحة	3,676 كتابًا، أيلول 2026
كلفة استخراج الغراف للمقطع	0.00035–0.00045 دولار (نحو 1,700 وحدة دخلًا و 1,100 خرجًا)	5,947 كتابًا معالجًا، تشرين الثاني 2026
زمن الاستعلام الهجين	أجزاء من الثانية قبل إعادة الترتيب	المجموعة الحية

تستحق ثلاث ملاحظات منهجية التصريح. الأولى أن دقة العزو تُقاس بمضاهاة المقطع المسترجع بصفحته في موقع المكتبة نفسها، وتُعاد معايرتها أسبوعيًا على عينة متجددة بتنبيه آلي عند الانحراف، على نهج الفحص بالعينات (ISO 2859-1) لا الفحص مرة واحدة عند الإضافة. الثانية أن الكلفة المدفوعة الوحيدة في هضم الكتاب هي التضمين، أما استخراج الغراف فهو البند الأكبر بفارق كبير (نحو 95% من كلفة إكمال المكتبة)، وهو ما يحكم ترتيب الأولويات في القسم التالي. الثالثة أن المكنز يخدم منذ تشرين الأول 2026 مستخدمين باحثين ومؤسسات في الإنتاج، وتقارير الاستخدام (النداءات بنوعها ووحدات إعادة الترتيب) تُقاس يوميًا، وهي مادة دراسة لاحقة عن أنماط السؤال البحثي في التراث.

8. الحدود والأعمال القادمة والإتاحة المفتوحة

حدود المنهج. المكنز يسترجع ما هُضم، فالكتاب الغائب غائب عن الجواب وإن كان أصل المسألة، ولذلك يُعلن نطاق التغطية على الموقع بالفنون والعدد. والتمثيل المتجهي يقرّب المعنى ولا يفهم السياق الأصولي للعبارة، فقد يجاور نصًا في مقام النقل بنص في مقام الردّ؛ وهنا تجيء قيمة الغراف الذي يميّز «يستند إلى» من «يعارض»، وهو بعدُ في أوله. والكتب المصورة تُقرأ آليًا عند الاستدعاء وتبقى دون النصية المحققة دقة. والنموذج اللغوي الذي يدير الأدوات من جهة المستخدم خارج سيطرة المكنز، فضمان المكنز يقف عند حدود ما تعيده أدواته: نص وفي بعزو صحيح.

الأعمال القادمة. (1) إكمال هضم المكتبة: نحو 11,000 مصنّف تضيف نحو عشرين مليون مقطع، بالنموذج نفسه للتضمين. (2) إكمال الغراف للمدونة كلها (نحو 32 مليون مقطع)، وهو ما يستدعي نقل الاستخراج إلى بنية حوسبة أرخص أو سيادية، بشرط واحد لا يُتجاوز: أن يساوي النموذج البديل القائم على المجموعة الذهبية أو يفوقه. (3) طبقات مخطط العلاقات للفنون التي لم تُستكمل (الرجال والفهارس). (4) مطابقة أسماء الكتب والمؤلفين عند الاستشهاد بسجل ألقاب وطبعات مبني من بيانات المكتبة، لأن أسماء التراث بيئة شديدة السيولة لا تضبطها المطابقة النصية.

الإتاحة المفتوحة. تلتزم الشركة بإخراج ثلاثة أشياء تحت رخص مفتوحة على Hugging Face و Zenodo:

1. **الغراف الببليوغرافي لمكتبة أهل البيت:** الأعلام والكتب والرواة والمفاهيم والعلاقات المصرّح بها بينها، مع مؤشرات الدليل (الكتاب والجزء والصفحة) وبلا نصوص الكتب.
2. **معيار مفتوح لاستخراج الكيانات والعلاقات من النثر العلمي العربي الكلاسيكي:** المجموعة الذهبية (150+ مقطعًا موسومًا عبر الفنون) ومخطط العلاقات وأداة القياس ونتائج النماذج المفتوحة عليها.
3. **هذه الورقة** بمعرّف DOI، وترجمتها الإنجليزية.

وصيغة الاستشهاد بالمكنز في الأعمال العلمية: صادقون للتقنية، «مكنز محيط (Moheet Corpus)»، الإصدار وتاريخ الاستعلام، moheet-ablibrary.net — مع العزو إلى الكتاب المطبوع نفسه في كل نص مسترجع، لأن المكنز وسيلة وصول لا مصدرًا.

المراجع

1. Ji, Ziwei, et al. "Survey of Hallucination in Natural Language Generation." *ACM Computing Surveys*, vol. 55, no. 12, 2023, pp. 1–38.
2. شادي علي. «مكنز محيط: نقلة في هيكلية المعارف الإسلامية الشيعية». لندن: صادقون للتقنية، أيلول 2026. https://moheet-ablibrary.net/docs/moheet-paper.html
3. Romanov, Maxim, and Masoumeh Seydi. "OpenITI: A Machine-Readable Corpus of Islamicate Texts." Zenodo, 2019–. https://openiti.org ؛ ومشروع KITAB (معهد الدراسات الإسلامية، جامعة الآغا خان، لندن). https://kitab-project.org

Kamran, Amna Binte, Bakhtawar Abro, and Amna Basharat. .4 "SemanticHadith: An Ontology-Driven Knowledge Graph for the Hadith Corpus." *Journal of Web Semantics*, vol. 78, 2023, Article .100797

Safe, M., et al. "Constructing a Knowledge Graph for Hadith .5 Narrators." *Journal of King Saud University – Computer and .Information Sciences*, vol. 34, no. 6, 2022, pp. 3385–3396

Qdrant. *Qdrant Vector Database Documentation*. .6 https://qdrant.tech/documentation

Reimers, Nils, and Iryna Gurevych. "Sentence-BERT: Sentence .7 Embeddings Using Siamese BERT-Networks." *Proceedings of .EMNLP-IJCNLP 2019*, ACL, 2019, pp. 3982–3992

Karpukhin, Vladimir, et al. "Dense Passage Retrieval for Open- .8 Domain Question Answering." *Proceedings of EMNLP 2020*, ACL, .2020, pp. 6769–6781

Cormack, Gordon V., Charles L. A. Clarke, and Stefan Büttcher. .9 "Reciprocal Rank Fusion Outperforms Condorcet and Individual Rank Learning Methods." *Proceedings of SIGIR 2009*, ACM, 2009, pp. .758–759

Malkov, Yu. A., and D. A. Yashunin. "Efficient and Robust .10 Approximate Nearest Neighbor Search Using Hierarchical Navigable Small World Graphs." *IEEE Transactions on Pattern Analysis and .Machine Intelligence*, vol. 42, no. 4, 2020, pp. 824–836

Hogan, Aidan, et al. "Knowledge Graphs." *ACM Computing* .11 .*Surveys*, vol. 54, no. 4, 2021, Article 71

Traag, V. A., L. Waltman, and N. J. van Eck. "From Louvain to .12 Leiden: Guaranteeing Well-Connected Communities." *Scientific .Reports*, vol. 9, 2019, Article 5233

Edge, Darren, et al. "From Local to Global: A Graph RAG Approach .13 .to Query-Focused Summarization." arXiv:2404.16130, 2024

Model Context Protocol. *Specification*. .14 https://modelcontextprotocol.io

Lewis, Patrick, et al. "Retrieval-Augmented Generation for .15 Knowledge-Intensive NLP Tasks." *Advances in Neural Information .Processing Systems 33*, 2020

ISO 2859-1:1999, *Sampling procedures for inspection by attributes*; ISO/IEC 25012:2008, *Data quality model*; IFLA, *Library Reference Model (LRM)*, 2017. [16]