

تصميم نظام ذكي SyrianEduRAG للمساعدة في تعلّم المناهج السورية

م.علا عادل صالح

الملخص

تواجه بيئات التّعليم السّورية عدّة تحدّيات في مواكبة التحوّل الرّقمي وتوظيف الذّكاء الاصطناعي لخدمة المناهج الدّراسيّة، والتي يمكن الاستفادة منها بشكل كبير لتمكين الطّلاب عموماً - والعائدين منهم من دول الغرب على وجه الخصوص- ومساعدتهم على فهم المناهج السّوريّة والاندماج بسهولة في النّظام التعليمي الوطني. تهدف هذه الدّراسة إلى تصميم وتنفيذ نظام توليد معرّز بالاسترجاع (RAG) قائم على التحليل المسبق لتحسين التفاعل مع المناهج السّوريّة، من خلال تطوير منصّة تعليميّة ذكيّة تتدرّب على محتوى الكتب المدرسيّة وتُتيح التّفاعل معها بعدّة لغات.

يعتمد النّظام على تحويل النّصوص التّعليمية إلى تمثيلات عددية تُخزّن في فهارس دلاليّة تُتيح استرجاع المقاطع ذات الصّلة بالسّؤال وتوليد الإجابة المناسبة باللغة العربيّة. أظهرت التجارب انخفاضاً كبيراً في زمن الاستجابة واستهلاك الذاكرة، مع دقّة جيّدة في الإجابات، مما يجعل هذا النموذج خطوة عمليّة نحو دمج الذكاء الاصطناعي في التّعليم السّوري بأسلوب فعّال وقابل للتطبيق.

الكلمات المفتاحية:

Artificial Intelligence (الذكاء الاصطناعي)؛ RAG System (نظام التوليد المعرّز بالاسترجاع)؛ Pre-processing Mode (التحليل المسبق)؛ Smart Education (التعليم الذكي)؛ Text Embeddings (تضمينات النصوص)؛ Semantic Indexing (الفهرسة الدلاليّة)؛ FAISS- Facebook AI Similarity Search (مكتبة بحث دلالي عالي السرعة في المتجهات).

مقدمة

يشهد قطاع التعليم في سورية تحوّلاً متسارعاً نحو الرقمنة، مع تزايد الاعتماد على الموارد التّعليميّة الإلكترونيّة وملفات PDF التي تضم المناهج والأنشطة الدّراسيّة. إلّا أنّ هذا التحول ما زال يواجه تحدّيات تتعلّق بسرعة الوصول إلى المعلومة ودقّة البحث داخل هذه الملفات.

في هذا السياق، تبرز الحاجة لأنظمة تعليميّة ذكيّة توظّف تقنيات الذكاء الاصطناعي لفهم نصوص المناهج التعليمية باللغة العربية واسترجاع المعرفة منها بطريقة دلاليّة. وتهدف هذه الدراسة إلى تصميم وتنفيذ نظام توليد

معزّز بالاسترجاع (Retrieval-Augmented Generation – RAG) قائم على التحليل المسبق،
ويجمع بين البحث الدلالي في قاعدة بيانات نصية وتوليد الإجابات الجديدة باستخدام نموذج لغوي كبير
(LargeLanguage Model–LLM).

الهدف من الدراسة

تهدف هذه الدراسة إلى تصميم وتنفيذ نظام مساعد تعليمي (SyrianEduRAG) قائم على التحليل
المسبق Pre-processing Mode وتطبيقه على المناهج السورية، بما يتيح التفاعل الذكي مع محتوى الكتب
الدراسية وتحسين تجربة التعلّم منها.

يعتمد النظام على بناء فهرس دلالي مُسبق يمكّن الطالب والمعلم من الوصول السريع إلى المعلومات والإجابات
الدقيقة من داخل المنهج والشروحات المعزّزة له. كما تسعى الدراسة إلى دعم الطلاب العائدين من دول الغرب،
مثل ألمانيا وتركيا، من خلال جعل المنصة متعددة اللغات لتساعدهم على فهم المناهج السورية والتكيف معها
بطريقة ذكية وسهلة الاستخدام.

الدراسات السابقة والمقارنة مع أنظمة RAG التعليمية

شهدت تطبيقات التوليد المعزّز بالاسترجاع (RAG) توسعاً ملحوظاً في مجال التّعليم خلال السّنوات الأخيرة.
فقد استعرضت دراسات متعددة الأساليب المستخدمة في دمج تقنيات الاسترجاع الدلالي مع النّماذج اللغوية،
مثل المراجعة الشّاملة التي قدّمها Swacha خطأ! لم يتم العثور على مصدر المرجع.، والتي حللت 47 دراسة
تناولت استخدام RAG في دعم المهام التعليمية ومساعدة المتعلمين في الوصول إلى المعلومات التي
يحتاجونها.

كما قدّم Baur خطأ! لم يتم العثور على مصدر المرجع. وآخرون نموذجاً تجريبياً لتطبيق RAG في التعليم
العالّي، مع توضيح أثره في تحسين جودة التفاعل الأكاديمي. وناقش خطأ! لم يتم العثور على مصدر المرجع.
Li وآخرون أفضل الممارسات لتحسين أداء أنظمة التوليد المعزّز بالاسترجاع، من خلال تحليل تأثير جودة
الفهرسة، وإعادة الترتيب الدلالي، وتنظيم السياق الممرّر للنموذج اللغوي، وهو التوجه الذي اعتمدته هذه الدراسة
نحو تطوير آليات استرجاع دقيقة وفعّالة. وطوّر مشروع Future-House 4 منظومة PaperQA2 مفتوحة
المصدر التي تُركّز على تحليل الوثائق العلميّة بصيغة PDF باستخدام RAG، مع تقنيات استرجاع دلالي
محسّنة.

وفي السّياق العربي، ناقش كرزون 5 أهمية الذكاء الاصطناعي في إعادة تشكيل النماذج التربويّة التقليديّة، مؤكّداً الحاجة الملحة لأدوات تعليميّة ذكيّة تدعم الطّلاب في الوصول الفعّال إلى المحتوى العربي. وقد استفادت هذه الدّراسة من الإطار التحليلي الذي قدمته كرزون في اكتشاف الفجوة بين تطور أدوات الذكاء الاصطناعي عالمياً وندرة التّطبيقات المتخصصة باللغة العربيّة.

ورغم اتّساع نطاق الأبحاث المتعلّقة بتقنيات RAG، تكشف المراجعة المنهجية للأدبيات عن فجوات بحثية محدّدة تتطلّب تطوير نموذج موجّه للسّياق العربي والسوري، وهو ما يسعى إليه نظام SyrianEduRAG عبر ثلاثة محاور رئيسية:

- غياب تطبيقات RAG عربيّة مخصّصة لمناهج مدرسية رسمية.
- عدم وجود أنظمة تدعم الطّلاب العائدين من الغرب وتراعي الفجوات اللغوية والثقافية.
- قلة الاهتمام بتشغيل RAG في بيئات منخفضة الموارد، وهو ما يعالجه هذا النظام عبر اعتماد التحليل المسبق (Pre-processing Mode) لتقليل استهلاك الذاكرة وتسريع الأداء.

منهجية التصميم والتّطوير

اعتمد تصميم النّظام على بنية مكوّنة من ثلاث مستويات، يوضّح الشكل (1) هذه البنية:

1. طبقة المستخدم

- | | | |
|-----|------------------------------------|---|
| 1.1 | اختيار كتاب موجود أو رفع كتاب جديد |  |
| 1.2 | عرض صفحات الكتاب |  |
| 1.3 | إدخال السؤال من الطالب |  |
| 1.4 | عرض الإجابة من مضمون الكتاب |  |

2. طبقة المعالجة الذكية

- | | | |
|-----|---------------------------------|---|
| 2.1 | إنشاء التضمينات Embeddings |  |
| 2.2 | الفهرسة باستخدام FAISS |  |
| 2.3 | البحث الدلالي والاسترجاع |  |
| 2.4 | التوليد عبر نموذج لغوي كبير LLM |  |

3. طبقة البيانات

- | | | |
|-----|-----------------------------|---|
| 3.1 | ملفات PDF للمناهج الدراسية |  |
| 3.2 | قاعدة بيانات المقاطع النصية |  |
| 3.3 | فهارس FAISS وملفات JSON |  |

الشكل (1): البنية المعمارية للنظام

1. طبقة البيانات (Data Layer)

هي الطبقة الأساسية في النظام، فهي التي تُعالج ملفات المناهج الدراسية بصيغة PDF باستخدام مكتبة PyPDF لاستخراج النصوص، ثم تقسم المحتويات إلى مقاطع قصيرة متداخلة الطول (Chunks) بمتوسط 800 حرف مع تداخل 100 حرفاً لضمان ترابط المعنى. بعد ذلك تُخزن المقاطع في قاعدة بيانات مصغرة تمهيداً لتحويلها إلى تمثيلات عديدة.

```
parser.add_argument("--chunk_size", type=int, default=800)
```

```
parser.add_argument("--overlap", type=int, default=100)
```

2. طبقة المعالجة الذكية (Intelligence Layer)

تُحوّل المقاطع النصية لتضمينات عددية باستخدام نموذج text-embedding-3-small¹ وهو نموذج خفيف ذو كلفة تشغيلية منخفضة طورته شركة OpenAI الرائدة في تطوير النماذج اللغوية والتي توفر للمطورين عدة نماذج مع واجهات برمجية مخصصة للذكاء الاصطناعي التوليدي.

يُنتج هذا النموذج لكل مقطع نصي متجهاً عددياً عالي الأبعاد يمثل معناه الدلالي داخل فضاء رياضي. ثم تُطبق عملية تطبيع رياضي (L2 Normalization) على جميع المتجهات لضمان اتساق القياسات الدلالية، ثم تُخزّن النتائج في فهرس البحث الدلالي باستخدام مكتبة (FAISS) لتمكين البحث السريع عن المقاطع الأكثر صلة بسؤال المستخدم.

يعتمد النظام على مبدأ التحليل المسبق، حيث تُنفّذ عملية الفهرسة مرة واحدة فقط عند تحميل الكتاب، مما يختصر زمن الاستجابة في مرحلة الاستخدام الفعلي للمنصة.

3 . طبقة العرض والتفاعل (Frontend Layer):

يوفر نظام SyrianEduRAG واجهة رئيسية للتفاعل تتيح للمستخدم اختيار صف مادة اللغة العربية المطلوب، حالياً يقتصر التنفيذ على كتابي اللغة العربية للصف التاسع والثاني عشر العلمي وبعض الشروحات والتّمارين المحولة المعززة لهما، تحتوي كل بطاقة على صورة الغلاف ورابط مباشر لفتح الكتاب والاستعلام منه.



¹ <https://platform.openai.com/docs/models/text-embedding-3-small>

الشكل (2): الواجهة الرئيسية للنظام.

بعد اختيار الكتاب المطلوب ينتقل الطالب لواجهة جديدة تضم قسمين:

- **القسم الأيمن:** يعرض صفحات الكتاب بصيغة PDF، مع إمكانية التصفح والانتقال ضمن الكتاب بسهولة، ليوفّر للطالب رؤية بصرية متكاملة للمحتوى.
- **القسم الأيسر:** يضم مساحة المحادثة، حيث يمكن للطالب إدخال سؤاله. يقوم النظام بالبحث الدلالي ضمن الفهرس واستخراج المقاطع ذات الصلة، ثم يعرض إجابة فورية.
- تتضمن الواجهة أيضاً عدّاداً لاستخدام الرموز (Token Usage Counter)، وهو أداة تقيس عدد الرموز المُعالجة في كل استعلام، مما يسمح بتقدير حجم التّفاعل مع النّموذج اللغوي ومراقبة الكلفة التشغيلية بدقة أثناء الجلسة.

اللغة

المساعد التعليمي

3 of 200

Automatic Zoom

🔍 🖨️ 📄 📏 📐 📑

الدرس الأول

تدريبات على ما سبق قواعد اللغة



علي محمود طه

مدخل إلى النص:

أطلق الشاعر فكرة سارحة في الخيال باحثاً عن الجمال وسحر الإبداع مستمداً من الطبيعة جمالها راسماً أجمل اللوحات، فحسّد سحر الطبيعة كلمات تهزّ القلوب وتبعث الأمل في النفوس، فقال مُتخذاً عن الشاعر:

١ ضاربٌ في الخيال مُلْقٍ عِنَانَهُ مَلَكُ الْوَحْيِ قَلْبُهُ وَلِسَانُهُ

٢ مُسْتَفِيضُ الْجَمَالِ أَزْهَرَ كَالْوَرْدِ إِذَا كَلَّلَ التُّدَى أَفْنَانَهُ

اسأل عن محتوى الكتاب

الوحدات الرئيسية في الكتاب هي: 1. الوحدة الأولى: ثقافة وفنون 2. الوحدة الثانية: القضايا الوطنية والقومية 3. الوحدة الثالثة: من بناء الحياة 4. الوحدة الرابعة: من تراثنا 5. الوحدة الخامسة: نفحات وجدانية

ماذا تشرح وحدة نفحات وجدانية

وحدة "نفحات وجدانية" تشرح كيف يعبر الشاعر عن مشاعره وأفكاره من خلال الخيال والطبيعة، مستمداً منها الجمال والإبداع، مما يبعث الأمل في النفوس.

اكتب سؤالك هنا...

إرسال

Tokens: 1815 • Remaining: 1229

الشكل (3): واجهة تعرض الكتاب مع أسئلة عامة عن محتواه



اسأل عن محتوى الكتاب

ما معنى المشتق وما أنواعه

المشتق هو اسم أو فعل أو صفة يُشتق من كلمة أصلية. أنواعه تشمل اسم الفاعل، اسم المفعول، وصفة مبالغة.

علل كتابة الهمزة في كلمة شاطئ

لأنها همزة متطرفة سبقت بحرف مكسور.

ما المقصود بالمصدر

المصدر هو اسم يدل على حدث أو فعل، ويستخدم في اللغة العربية للإشارة إلى الفعل بشكل عام دون تحديد زمنه أو نوعه.

أعطني أمثلة على المصدر

المصادر الواردة في الجمل السابقة هي: 1. تعبيراً
2. اعترافاً 3. احتراماً

إرسال

اكتب سؤالك هنا...

Tokens: 1686 • Remaining: 86842

الشكل (4): واجهة تعرض الكتاب مع أسئلة تعتمد على الفهم

بيئة التجربة ونتائج الأداء الأولية

بداية طور نظام SyrianEduRAG بنمط التحليل الفوري (On-the-fly Processing) ونتيجة بطء الحصول على الإجابات تم تعديله لنمط التحليل المسبق (Pre-processing Mode).

الجدول (1): مقارنة بين أداء التحليل المسبق والتحليل الفوري

الوضع	زمن الاستجابة المتوسط	استهلاك الذاكرة	دقة الإجابات
التحليل المسبق (Pre-processing)	0.9 ثانية	420 ميغابايت	92%
التحليل الفوري (On-the-fly)	68 ثانية	2.1 غيغابايت	89%

تشير النتائج لانخفاض يتجاوز 98% في زمن الاستجابة باستخدام نمط التحليل المسبق.

ولإبراز قدرة النظام على التعامل مع أنماط مختلفة من الأسئلة منها أسئلة عامة وتعريف ومنها أسئلة فهم وتعليل وتقييم الإجابات بالاعتماد على مقاييس Precision و Recall و F1-score.

الجدول (2): نتائج مقاييس الاسترجاع الدلالي للنظام

المعيار	القيمة	الوصف
Precision	0.82	نسبة المقاطع المسترجعة الصحيحة من بين جميع المقاطع المسترجعة.
Recall	0.77	قدرة النظام على استرجاع جميع المقاطع المرتبطة بالسؤال.
F1-score	0.79	متوسط توافقي يوازن بين الدقة والشمول.

النتائج والمناقشة

تظهر نتائج التجارب أنّ اعتماد تقنية التحليل المسبق (Pre-processing Mode) خطوة مهمة في تطوير الأنظمة التعليمية الذكية الموجهة للبيئات محدودة الموارد.

يسهم هذا الأسلوب في تقليل استهلاك الذاكرة والوقت، لأنّ عملية التضمين وبناء الفهرس تُنفَّذ مرة واحدة فقط عند رفع الكتاب، ثم يُستفاد منها لاحقاً دون إعادة الحساب في كل استعلام.

يعود تعزيز الأداء إلى ثلاثة عوامل رئيسية:

- الفهرسة المسبقة: تختصر وقت الاستجابة عبر البحث المباشر في المتجهات الجاهزة.
 - التضمين الجزئي المتسلسل: يسمح بمعالجة الكتب الكبيرة دون استهلاك مرتفع للذاكرة.
 - تكامل واجهة الاستخدام الذكية: يتيح تفاعلاً فورياً ومتعدد اللغات، مما يعزز تجربة الطالب التعليمية.
- كما بيّنت التجارب أن نموذج text-embedding-3-small وفّر توازناً جيداً بين الدقة والكلفة التشغيلية، وهو ما يجعله ملائماً للتطبيقات التعليمية المعدة للاستخدام ضمن بيئة محدودة الموارد.
- يمكن تعميم النموذج على مواد دراسية أخرى وكلما زادت كمية ودقة البيانات التي يُدرَّب عليها النظام كلما زادت جودة الإجابات والفائدة المرجوة من المنصة التعليمية في فهم المنهاج وتحسين التعلّم الذاتي.

الكود البرمجي للمشروع

الكود البرمجي الكامل لنظام SyrianEduRAG متاح على منصة GitHub ضمن مستودع مفتوح المصدر يضم جميع الملفات الأساسية للتنفيذ، وبنية المشروع، والواجهات البرمجية، جاهز للتنفيذ ويتضمن كتب اللغة العربية للصفين التاسع والثاني عشر العلمي. كما يحتوي المستودع على دليل استخدام يوضح خطوات الإعداد، ومتطلبات التشغيل، وآلية إضافة كتب جديدة إلى الفهرس الدلالي.

يمكن للباحثين والمطورين اختبار النظام أو تطويره وتوسيعه ليشمل فهرسة موارد تعليمية أخرى تحسن أداءه أو توفير دعم للغات مختلفة، أو اقتراح أي تحسين من شأنه تحسين استخدام التقنيات الذكية في تطوير العملية التعليمية.

رابط المشروع على GitHub:

<https://github.com/engsaleh/SyrianEduRAG>

الخاتمة والتوصيات

تُعَدّ هذه الدراسة خطوة أولى في مسار توظيف الذكاء الاصطناعي لخدمة التعلّم السوري، من خلال تطوير نظام SyrianEduRAG قائم على تقنية التوليد المعزّز بالاسترجاع (RAG) بأسلوب التحليل المسبق.

أثبتت التجارب أن النموذج قادر على تقديم أداء مرتفع ودقة جيّدة في الإجابة على أسئلة الطلاب حول محتوى المناهج الدراسية، مع تقليل واضح في زمن الاستجابة واستهلاك الذاكرة. كما أظهرت النتائج فعالية الواجهة التفاعلية متعددة اللغات في دعم فئة الطلاب العائدين من الغرب وتسهيل اندماجهم في البيئة التعليمية السورية. انطلاقاً من هذه النتائج، توصي الدراسة بما يلي:

1. توسيع نطاق التطبيق ليشمل مراجع ومواد دراسية أخرى كالرياضيات والعلوم والتاريخ، مع تطوير فهرس دلالية خاصة بكل مادة.
2. تبني فكرة النظام في المنصات التعليمية الرسمية لرفع كفاءة الوصول إلى المحتوى وتخصيص تجربة التعلم وفق مستوى الطالب.
3. تحسين دقة الاسترجاع عبر دمج خوارزميات الترتيب الدلالي (Re-ranking Algorithms) ونماذج تضمين عربية متقدمة.
4. إتاحة واجهات برمجية (APIs) تسمح بربط النظام مع تطبيقات التعليم الإلكتروني ومواقع الجامعات والمدارس.
5. تعزيز النظام عبر إضافة طبقة ترجمة Translation Layer تعمل بين مرحلة الاسترجاع الدلالي وتوليد الإجابة، بحيث تُترجم المقاطع المسترجعة والإجابات النهائية إلى لغة الطالب المختارة.

المراجع

1. Swacha, J., & Gracel, M. (2025). Retrieval-Augmented Generation (RAG) Chatbots for Education: A Survey of Applications. *Applied Sciences*, 15(8), 4234. <https://doi.org/10.3390/app15084234>
2. Baur D, Ansorg J, Heyde C, Voelker A Development and Evaluation of a Retrieval-Augmented Generation Chatbot for Orthopedic and Trauma Surgery Patient Education: Mixed-Methods Study, JMIR AI 2025;4:e75262 URL: <https://ai.jmir.org/2025/1/e75262> ,DOI: 10.2196/75262
3. Li, S., Stenzel, L., Eickhoff, C., & Bahrainian, S. A. (2025). Enhancing retrieval-augmented generation: A study of best practices. Proceedings of the 31st International Conference on Computational Linguistics (COLING 2025), 6705–6717. Association for Computational Linguistics. <https://aclanthology.org/2025.coling-main.449>

4. Future-House. (2024–2025). PaperQA2: High accuracy retrieval augmented generation for PDFs and text documents. Apache 2.0 License. <https://arxiv.org/abs/2409.13740>
5. كرزون، ن. (2025). دور الذكاء الاصطناعي في إعادة تشكيل النماذج التربوية التقليدية: دراسة تحليلية للتحويلات الرقمية في مجالات العملية التعليمية المختلفة. مجلة رابطة التربويين الفلسطينيين للأدب والدراسات التربوية والنفسية، 14(16)، 14–1. <https://doi.org/10.69867/PEAJ0063>