

توليد الأسئلة متعددة الخيارات باللغة العربية من المراجع العلمية الأجنبية

م. محمد سليمان^{1*}, د. عمار جوخدار²^{1*} طالب ماجستير، جامعة دمشق، ذكاء اصطناعي² أ. د.، جامعة دمشق، ذكاء اصطناعي

الملخص

لا تزال اللغة الإنكليزية اللغة الأغنى من حيث عدد المراجع العلمية الحديثة والمستخدم في معظم الاختصاصات ومنها الطب والهندسة. ويعتمد العديد من المدرسين الجامعيين على المراجع المكتوبة باللغة الإنكليزية في تطوير المناهج التعليمية ووضع الأسئلة الامتحانية، وفي المقابل، يعاني الطلاب الغير ناطقين باللغة الإنكليزية من صعوبة فهم وتفسير هذه المراجع. نهدف في هذا البحث إلى مساعدة الطلاب العرب على تنمية مهاراتهم واختبار مدى فهمهم للمادة العلمية من خلال توليد الأسئلة متعددة الخيارات باللغة العربية من المرجع العلمي الأجنبي. وكذلك يمكن للمدرسين الاستفادة من الأسئلة الناتجة عند وضع الأسئلة الامتحانية. يتم ذلك من خلال استخراج نموذج معرفة من المرجع الأجنبي ثم مطابقة النموذج مع قوالب أسئلة معدة مسبقاً باللغة العربية، يتم استخدام نموذج المعرفة أيضاً للحصول على الخيارات الخاطئة من داخل المرجع العلمي. قمنا بإجراء التجربة على مرجع علمي في المجال الطبي، تم توليد أكثر من 3200 سؤال متعدد الخيارات باللغة العربية خلال 7.9 ثانية، بلغ عدد الأسئلة ذات القصة السريرية 114 سؤال، وبحسب تقييم الخبراء لعينة عشوائية من الأسئلة، فإن 80% من الأسئلة كانت صحيحة وذات درجات صعوبة مقبولة، وتم تحقيق قيمة 0.81 على مقياس Fleiss' kappa لتوافق الخبراء، مما يعني أن الأسئلة الناتجة يمكن استخدامها في عملية تعليم الطلاب وتقييم مهاراتهم العلمية.

تاريخ الايداع

تاريخ القبول



حقوق النشر: جامعة دمشق -

سورية، يحتفظ المؤلفون بحقوق

النشر بموجب CC BY-NC-SA

الكلمات المفتاحية: الأسئلة متعددة الخيارات، معالجة اللغات الطبيعية، شبكة معرفة، التعليم، الطب

Generating Arabic Multiple-Choice Questions from International Scientific References

Muhammad Sulaiman^{1*}, Dr. Ammar Joukhadar²

¹ Master degree student, Damascus university, Artificial Intelligence

² PhD, Prof., Damascus university, Artificial Intelligence

Abstract

Modern scientific references are mostly written in English in most of the fields, including medicine and engineering. Many university instructors rely on English references to develop educational curricula and formulate exam questions. In contrast, non-English-speaking students struggle to understand and interpret these references. In this research, we aim to assist Arab students in developing their skills and testing their understanding of the scientific material by generating Arabic multiple-choice questions from English scientific references. Additionally, teachers can benefit from the generated questions when preparing exam questions. This was achieved by, first, extracting a knowledge model from the English reference, then, passing Arabic question templates throughout the model to generate questions in Arabic. The knowledge model is also used to obtain incorrect options from within the scientific reference. We conducted the experiment on a scientific reference in the medical field, generating over 3,200 multiple-choice questions in Arabic in 7.9 seconds. The number of medical case-based questions was 114. According to expert evaluation of a random sample of the questions, 80% of the questions were correct and had acceptable difficulty levels. A value of 0.81 was achieved on the Fleiss' kappa scale for expert agreement, indicating that the generated questions can be used in the process of teaching and assessing students' scientific skills.

Keywords: Multiple Choice Questions, NLP, Knowledge graph, Education, Medicine

Received:

Accepted:



Copyright: Damascus University- Syria, The authors retain the copyright under a CC BY- NC-SA

1 . المقدمة (Introduction)

(translation), وغير ذلك من الخدمات التي تساعد الطالب على فهم المادة العلمية¹.

نهدف في هذا البحث إلى مساعدة الطلاب العرب عند تناولهم للمادة العلمية الموجودة في المرجع الأجنبي المكتوب باللغة الإنكليزية، وذلك من خلال تطوير آلية لتوليد الأسئلة متعددة الخيارات باللغة العربية من المرجع الأجنبي، مع تغطية جميع مواضيع الكتاب والتحكم بصعوبة الأسئلة الناتجة. وبالتالي، يمكن للطلاب اختبار مدى فهمه الشخصي للمادة العلمية، وكذلك، يمكن للمدرسين الاستفادة من الأسئلة الناتجة في تسريع عملية وضع الأسئلة الامتحانية. وفي النهاية، يمكن استخدام هذه الآلية ضمن تطبيق "الكتاب الذكي"، حيث يستطيع المستخدم استعراض مجموعة أسئلة متعددة الخيارات حسب الموضوع المرغوب، مع يتم ربط كل سؤال بمصدره في الكتاب.

بقية الورقة مقسمة كما يلي: نستعرض في الفقرة التالية الأبحاث ذات الصلة مع الإشارة إلى الثغرات الموجودة فيها بهدف تبرير العمل في هذا البحث، ثم نقوم في الفقرة الثالثة بشرح أهم المفاهيم النظرية ومنهجيات الحل المتبعة لحل المسألة بما يساعد في فهم الأساس النظري للحل المقترح والذي سنقوم باستعراضه في الفقرة الرابعة. نقوم في الفقرة الخامسة بوصف تجربة عملية تم إجراؤها على مرجع علمي في المجال الطبي، ثم نقوم بتقييم ومناقشة النتائج بشكل موضوعي في الفقرة السادسة.

2 . الدراسة المرجعية (Literature Review)

بدأ العمل على توليد الأسئلة آلياً منذ التسعينات (CH & Saha, 2020)، ولا يزال العمل في هذا المجال نشطاً. وبشكل عام، عملت معظم الدراسات التي تناولت مسألة "توليد الأسئلة من الوثائق التعليمية" على اللغة الإنكليزية، وذلك بسبب ضعف أدوات معالجة اللغات الطبيعية (NLP) في اللغات الأخرى

تعتبر المراجع العلمية الأجنبية بمثابة مصدر المعلومات الأكثر أهمية في مجال التعليم العالي، وخاصة في المجالات العلمية التخصصية مثل الطب والهندسة. ويعاني الطلاب الغير ناطقين باللغة الإنكليزية من صعوبة فهم هذه المراجع ودراساتها، فمن ناحية، يعتمد المدرسون الجامعيون على المراجع العلمية الأجنبية في عملية وضع الأسئلة الامتحانية وتطوير المناهج التعليمية، ومن ناحية أخرى، لا توجد حركة ترجمة للمراجع الأجنبية إلى بقية اللغات، خاصة اللغة العربية، ويعود ذلك لعدة أسباب، منها، سرعة تحديث المعلومات الموجودة في المراجع والكلفة العالية لعملية الترجمة.

على صعيد آخر، يعتبر طرح الأسئلة أمراً بالغ الأهمية في العملية التعليمية، فمن ناحية، يمكن للطلاب أن يعزز فهمه للمعلومة من خلال الإجابة على سؤال متعلق بها، ومن ناحية أخرى، يعتمد المدرسون على وضع أسئلة متنوعة الصعوبة بهدف اختبار الطلاب وتقييم مهاراتهم العلمية. وتعتبر الأسئلة متعددة الخيارات أحد أهم أنواع الأسئلة المعتمدة في المجال التعليمي، خاصة مع الزيادة الكبيرة في أعداد الطلاب وانتشار منصات التعلم عن بُعد.

توجد العديد من التطبيقات التي تستخدم تقنيات الذكاء الاصطناعي في العملية التعليمية، على سبيل المثال، ظهر مفهوم "الكتاب الذكي" (Intelligent textbook) كوصف للتطبيق (Application) الذي يقوم بعرض الكتاب العلمي للطلاب مع بعض الخدمات الذكية مثل خدمة الإجابة الآلية عن الأسئلة (question answering) لمساعدة الطالب على فهم النص العلمي، وخدمة توليد الأسئلة آلياً (question generation) بهدف تدريب الطالب واختبار فهمه للمادة التعليمية (Jiang et al., 2023)، وخدمات أخرى مثل تلخيص النصوص (text summarization)، التصفح الذكي (smart navigation)، الترجمة الآلية (machine

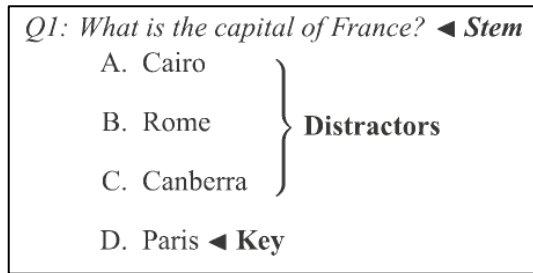
¹ مؤخراً، طوّرت غوغل (Google) إضافة (Google Scholar PDF Reader) الخاصة بتصفّح الانترنت، حيث تقوم هذه الإضافة بتحليل الورقة البحثية وتسمح للباحث بالنقر على الاستشهاد (citation) الموجودة ضمن النص لتظهر نافذة منبثقة تحتوي جميع المعلومات الخاصة بالبحث الذي تم النقر عليه، وبالتالي، المساعدة على ربط الأبحاث مع بعضها وتخفيف حجم العمل اليدوي الملقى على عاتق الباحث.

2 من 14

نستعرض في هذه الفقرة أهم المفاهيم ومنهجيات العمل المستخدمة في مسألة "توليد الأسئلة من الوثائق التعليمية" وذلك مما يساعد لاحقاً في فهم الحل المقترح.

1.3. السؤال متعدد الخيارات (MCQ)

يتألف السؤال متعدد الخيارات بشكل رئيسي من جسم السؤال، الخيارات الخاطئة والخيار الصحيح (شكل 1).



شكل 1. بنية السؤال متعدد الخيارات. (Leo et al., 2019)

ومن المهم ملاحظة أن عدد الخيارات الخاطئة يؤثر على عدد الأسئلة التي يمكن الحصول عليها. فمثلاً، إذا كانت لدينا أربعة خيارات خاطئة، يمكننا استعمال ثلاثة خيارات منها لتكوين سؤال واحد، ثم يمكننا استبدال أحد تلك الخيارات بالخيار الرابع، وبالتالي نحصل على سؤال جديد. ويمكننا حساب عدد الأسئلة التي يمكن الحصول عليها عند توفر n خيار خاطئ وفق (علاقة 1).

$$QuestionCount(Q, A, n) = C_n^3 = \frac{n!}{3!(n-3)!}$$

علاقة 1. حساب عدد الأسئلة متعددة الخيارات تبعاً لعدد الخيارات الخاطئة المتاحة.

على سبيل المثال، من أجل السؤال (Q) ذو الإجابة (A)، يمكننا توليد أربعة أسئلة مختلفة إذا كانت لدينا $n=4$ خيارات خاطئة، أما إذا كانت لدينا 5 خيارات خاطئة، فيمكننا حينئذٍ توليد 10 أسئلة.

2.3. التحكم بصعوبة السؤال (Difficulty Control)

يحتاج واضع الأسئلة إلى تأليف أسئلة تتناول مستويات صعوبة مختلفة، وذلك بهدف تمييز الطلاب حسب مهاراتهم العلمية، فالسؤال السهل يساعد على تمييز الطالب الأقل مهارة عن بقية الطلاب، والسؤال الصعب يساعد على تمييز الطالب

الطوب (والهندسة). علاوة على ذلك، يمكننا القول -ضمن حدود معرفتنا- بأنه لا يوجد أي بحث عمل على توليد الأسئلة باللغة العربية من النصوص أو الوثائق المكتوبة بلغة غير اللغة العربية.

مؤخراً، بدأ استخدام نماذج اللغة الكبيرة (LLM) في العديد من المجالات والتطبيقات، وقد حققت هذه النماذج نتائج جيدة، متفوقة على منهجيات العمل المختلفة. أما بالنسبة لاستخدام هذه النماذج في توليد الأسئلة باللغة العربية، فقد بينت العديد من الدراسات، مثل (Khondaker et al., 2023; Naous et al., 2024)، انخفاض أداء وجودة مخرجات هذه الأنظمة عند توليد النصوص باللغة العربية في مختلف المسائل، مثل مسألة الإجابة على الأسئلة (Question Answering) وتلخيص النصوص (Text summarization) وغيرها. وحتى بالنسبة لنماذج اللغة الكبيرة التي تمت معايرتها وتخصيصها للغة العربية (Arabic LLMs) مثل (Sengupta et Jais, 2023) و (Huang et al., 2023) AceGPT، فإنه لا تزال هناك الكثير من التحديات التي تواجه هذه النماذج، فعلى سبيل المثال، بين (Alghamdi et al., 2024) أن أداء هذه النماذج لم يتجاوز نسبة 60% في اختبار الإجابة على الأسئلة² المقدمة لها باللغة العربية!

نهدف في هذا العمل إلى تطوير آلية لتوليد الأسئلة باللغة العربية من المراجع العلمية الأجنبية في المجال الطبي، مع أخذ السمات البنوية (layout-features) للمرجع الأجنبي بعين الاعتبار واعتماد منهجية عمل تعتمد على قوالب الأسئلة (template-based)، مما يساعد على تجاوز عقبات اختلاف اللغة بين المرجع الأجنبي والأسئلة الناتجة باللغة العربية، وكذلك الحصول على أسئلة متعددة الخيارات سليمة لغوياً وذات درجات صعوبة متنوعة.

3. أرضية العمل (Background)

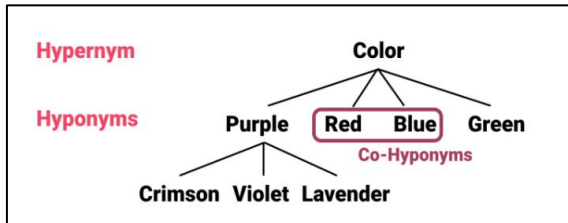
² تم في هذا الاختبار تحضير 516 سؤالاً متعدد الخيارات باللغة العربية، وطلب من نماذج Jais و AceGPT اختيار الجواب الصحيح، فكانت النتيجة أن استطاعت هذه النماذج الإجابة على حوالي 310 سؤال فقط بشكل صحيح، في حين فشلت بالإجابة على بقية الأسئلة (حوالي 200 سؤال).

السؤال أكثر صعوبة, وكلما كان التشابه أقل, كان السؤال أقل صعوبة.

3.3. توليد الخيارات الخاطئة (Distractor generation)

تعتبر مسألة توليد الخيارات الخاطئة أكثر صعوبة من مسألة توليد السؤال متعدد الخيارات بحد ذاته (Dijkstra, 2022), وذلك لأن اختيار خيارات خاطئة يستطيع جميع الطلاب استبعادها (بغض النظر عن مهاراتهم) يجعل السؤال كله بلا قيمة! ولذلك, تعتبر مسألة توليد الخيارات الخاطئة مسألة مستقلة بحد ذاتها, وتعمل بعض الأبحاث خصيصاً على هذه المسألة (Alhazmi et al., 2024), حيث يكون دخل النظام عبارة عن جسم السؤال (Stem) بالإضافة إلى الخيار الصحيح (Key), ويُطلب من النظام توليد الخيارات الخاطئة (Distractors), كما في (Han, 2022).⁴

توجد العديد من منهجيات اختيار الخيارات الخاطئة, منها ما يعتمد على السمات النحوية والقواميس (شكل 3)

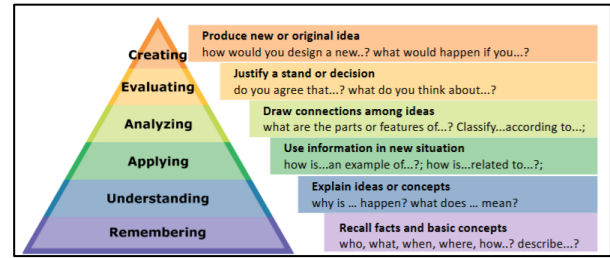


شكل 3. اختيار الجواب الخاطئ اعتماداً على السمات اللغوية والصرفية (كما في شبكة WordNet). مأخوذ من (Patil et al., 2022)

وبعض المنهجيات تعتمد على درجة التقارب بين المفاهيم في فضاء التمثيل الرقمي (مثل word2vec أو doc2vec), وبعض المنهجيات تعتمد على شبكات معرفة عامة (مثل ConceptNet) من أجل اقتراح مفاهيم مشابهة للخيار الصحيح.

الأكثر مهارة عن بقية الطلاب. وتعمل العديد من الأبحاث على تطوير منهجيات تسمح لواضع الأسئلة بالتحكم بصعوبة السؤال المتولد³.

توجد العديد من النظريات التي تهدف إلى تصنيف الأسئلة في المجال التعليمي بحسب درجة صعوبتها, وتعتبر هرمية بلوم (Bloom's Taxonomy) (Bloom et al., 1956), مرجعاً أساسياً لتصنيف الأسئلة, حيث ينتمي السؤال إلى أحد المستويات الستة في الهرم (شكل 2).



شكل 2. هرمية بلوم (Bloom's taxonomy) لتصنيف الأسئلة في المجال التعليمي. مأخوذ من (Pan, 2022)

بشكل عام, يمكن التحكم بصعوبة السؤال من خلال التحكم بشكل جسم السؤال (Stem). وبالتالي, يمكن تصنيف اسناد درجة صعوبة للسؤال تبعاً للمستوى الذي ينتمي إليه السؤال في هرمية بلوم (جدول 2). ويهدف البحث في هذا المجال للحصول على أنظمة توليد أسئلة تعليمية تستهدف المستويات العليا من هرمية بلوم.

جدول 2. تصنيف صعوبة الأسئلة حسب هرمية بلوم. مأخوذ من (Rahim et al., 2017)

Level 1	Knowledge	Easy
Level 2	Comprehension	
Level 3	Application	Medium
Level 4	Analysis	
Level 5	Synthesis	Hard
Level 6	Evaluation	

أيضاً, بالنسبة للأسئلة متعددة الخيارات, يمكن التحكم بدرجة صعوبة السؤال من خلال قياس درجة التشابه بين الخيار الصحيح والخيارات الخاطئة, فكلما كان التشابه أكبر, كان

³ تعمل بعض الأبحاث على مسألة قياس درجة صعوبة السؤال بهدف ترتيب الأسئلة الموجودة مسبقاً (ranking) في بنوك الأسئلة, وبالتالي إمكانية استرجاع أسئلة حسب مستوى الصعوبة المرغوب.

⁴ في دراسته لتوليد الخيارات الخاطئة بعدة لغات (عربية, إنكليزية, صينية, ألمانية), بين أن توليد الخيارات الخاطئة باللغة العربية كان الأصعب من بين جميع اللغات.

4.3. منهجيات الحل

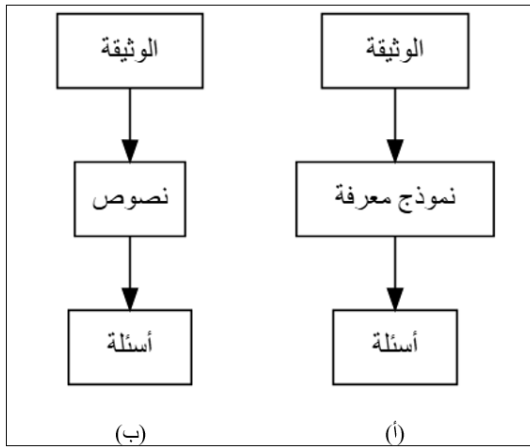
يمكن تقسيم منهجيات توليد الأسئلة من الوثائق إلى قسمين رئيسيين (شكل 4):

1. منهجيات تعتمد على استخراج نموذج معرفة

(knowledge model) من الوثيقة التعليمية: تقوم هذه المنهجيات على فكرة تحويل الوثيقة من بنية غير مهيكلة إلى بنية مهيكلة، ثم توليد الأسئلة من البنية المهيكلة، كما في (Dresscher et al., 2021).

وتعتمد فكرة تحويل الوثيقة التعليمية إلى نموذج معرفة على أساس أن الوثيقة التعليمية تحتوي على سمات بنيوية يمكن الاعتماد عليها لاستخراج نموذج المعرفة بشكل آلي، ومن أهم هذه السمات: فهرس المحتويات، بنية الصفحة، حجم ونوع الخط (العناوين الرئيسية والجزئية)، خصائص الجداول (أسماء الأعمدة وأعدادها)، الخ...

2. منهجيات تعتمد على استخراج النصوص من الوثيقة التعليمية، ثم توليد الأسئلة من النصوص المستخرجة باستخدام نماذج "توليد الأسئلة من النصوص". وتعتبر هذه المنهجية هي الأكثر انتشاراً، إلا أن هذه الطريقة هي الأقل كفاءة، وذلك لأن النصوص المستخرجة من الوثيقة تنتزع من سياقها، وبالتالي، فإن الأسئلة الناتجة لن تغطي جميع مستويات المعرفة الموجودة في الوثيقة التعليمية، إنما ستغطي مستوى المعرفة الموجود على صعيد الجملة والفقرة فقط، وبالتالي، يصبح من غير الممكن توليد أسئلة تتطلب مستوى عالياً من الفهم والادراك حتى يتمكن الطالب من الإجابة عليها.



شكل 4. منهجيات توليد الأسئلة من الوثائق التعليمية. (أ) منهجيات تعتمد على استخراج نموذج معرفة (knowledge model) من الوثيقة التعليمية، ثم تقوم بتوليد الأسئلة من نموذج المعرفة. (ب) منهجيات تعتمد على استخراج النصوص من الوثيقة التعليمية، ومن ثم تقوم بتوليد الأسئلة من النصوص المستخرجة.

4. الحل المقترح (Approach)

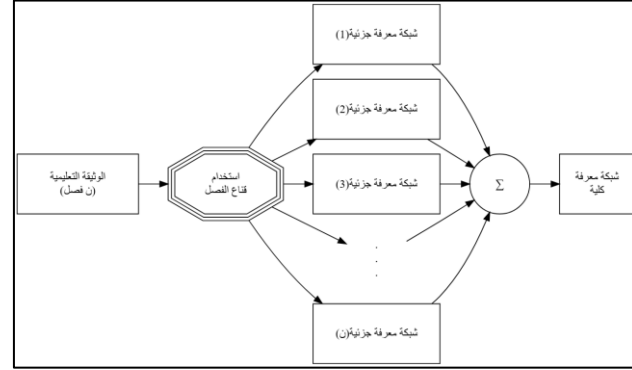
يتبع الحل المقترح المنهجية المعتمدة على استخراج نموذج معرفة من الوثيقة (المرجع العلمي) بحيث يتم استخدام نموذج المعرفة للحصول على الأسئلة متعددة الخيارات وكذلك لضمان الحصول على الخيارات الخاطئة من داخل محتوى الوثيقة. يتم الاعتماد بشكل أساسي على منهجية (layout-based) كما في العديد من الدراسات، مثل (Abdurahman et al., 2021; Dresscher et al., 2021)، وقد قمنا باختيار هذه المنهجية بناءً على ملاحظة كون كثير من المراجع العلمية تقوم بسرد المعلومات في فصول متشابهة من حيث البنية⁵، ويمكننا وصف هذه البنية بأنها شبه مهيكلة (semi-structured). وبالتالي، تقوم الفكرة على تحويل المرجع العلمي من بنية شبه مهيكلة إلى بنية مهيكلة تماماً، ونعتقد أن اختيار منهجية (layout-based) لمعالجة المراجع الشبه مهيكلة (semi-structured) سيساعد على:

1. استخراج نموذج معرفة تخصصي وغني مرتبط

بالوثيقة،

⁵ فمثلاً، في المجال الطبي، نلاحظ أن في المراجع الخاصة بعلم الأدوية، في كل فصل يتم التحدث عن دواء، يذكر الفصل تعريف الدواء، تركيبه، تأثيراته الجانبية، تداخلاته الدوائية مع الأدوية الأخرى، الخ... وكذلك بالنسبة لعلم الأمراض. وفي مجال الهندسة، يمكن ملاحظة ذلك في العديد من المراجع مثل ال Design Patterns

2. التقاط العلاقات بين المصطلحات في الوثيقة التعليمية نفسها.



شكل 6. استخدام "قناع الفصل" لاستخراج شبكات المعرفة من فصول المرجع العلمي

ثانياً، يتم اعداد قوالب أسئلة باللغة العربية مسبقاً بهدف تمريرها إلى نموذج المعرفة. وقد تم اختيار المنهجية المعتمدة على القوالب (template-based) لعدة أسباب:

1. يمكننا التحكم بصعوبة السؤال من خلال التحكم بشكل القالب

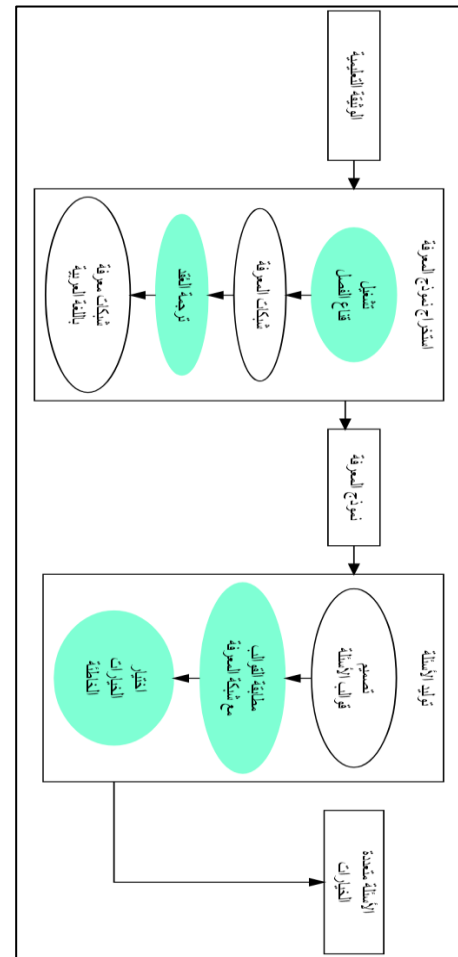
2. يمكننا -من خلال القالب- اظهار السؤال بشكل سليم لغوياً ومشابه للأسئلة التي يضعها الخبير.

وبشكل عام، تعتبر المنهجية المعتمدة على القوالب (template-based) من المنهجيات الأساسية والأكثر استخداماً في مسائل توليد الأسئلة في المجالات التعليمية (Willert & Thiemann, 2024).

ثالثاً، يتم استخدام قاموس تخصصي لترجمة العُقد (nodes) الموجودة في شبكة المعرفة من أجل الحصول على شبكة معرفة باللغة العربية. ومن خلال هذه الطريقة، يمكننا تجاوز التحديات المتعلقة بأنظمة الترجمة الآلية وانخفاض أدائها في حال استخدامها لترجمة السؤال التخصصي من اللغة الإنكليزية إلى اللغة العربية.

رابعاً، يتم تمرير قوالب الأسئلة على شبكة المعرفة بهدف الحصول على قيم للمتحويلات الموجودة في القوالب، ويتم اختيار الخيارات الخاطئة مع مراعاة درجة الصعوبة المرغوبة للسؤال،

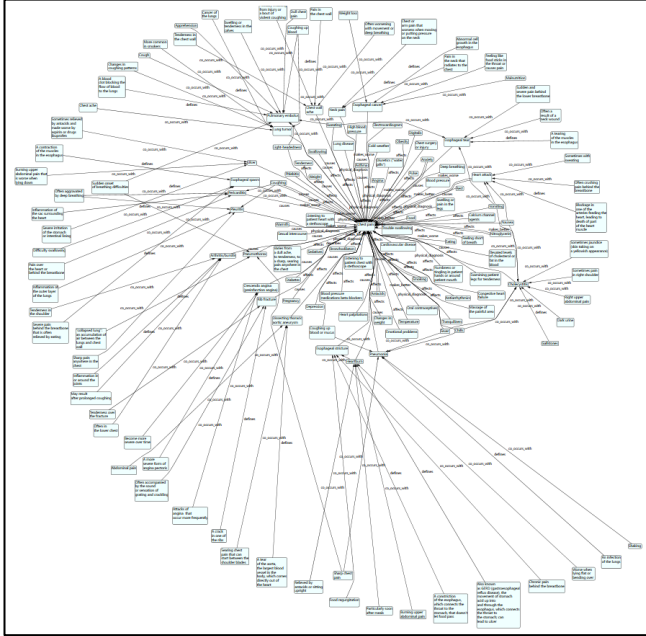
وعليه، فإننا نقترح منهجية شبه آلية (semi-automatic) (شكل 5) لاستخراج نموذج معرفة خاص بالمرجع العلمي، ومن ثم، يتم استخدام نموذج المعرفة مع قوالب أسئلة معدة مسبقاً باللغة العربية اللغات من أجل الحصول على عدد كبير من الأسئلة متعددة الخيارات مع التحكم بصعوبة السؤال. وفيما يلي، سنقوم بشرح خطوات العمل في الحل المقترح بشكل أكثر تفصيلاً.



شكل 5. مخطط الحل المقترح

أولاً، يتم تمرير الوثيقة على مكوّن "قناع الفصل" (شكل 6)، وهو عبارة عن برمجية تأخذ بعين الاعتبار بنية الوثيقة شبه المهيكلة من أجل استخراج البيانات من البنى المختلفة الموجودة في

جميع الأسئلة الممكن استخراجها بواسطة القالب (شكل 9). يتم اختيار الخيارات الخاطئة بحسب المسافة بينها وبين الخيار الصحيح بما يتوافق مع درجة الصعوبة المرغوبة.



شكل 8. شبكة المعرفة التي تم استخراجها من الفصل الخامس عشر (Chest pain) باستخدام قناع الفصل

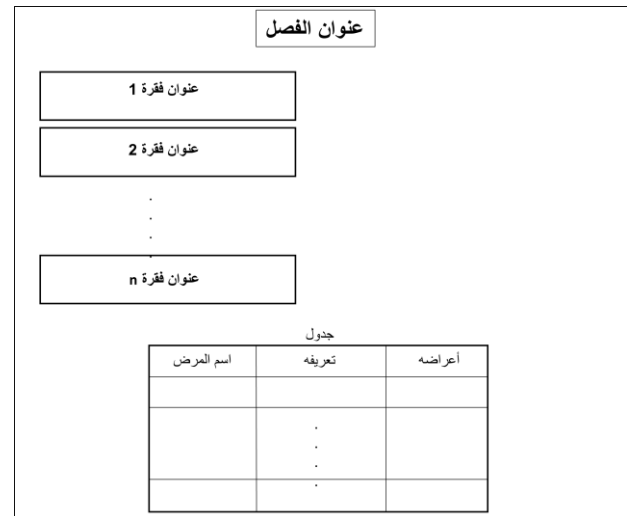
جدول 3. بعض القوالب المستخدمة في التجربة

مستوى الصعوبة	القالب (الخيارات)	القالب (جسم السؤال)	
سهل	[[Symptom]]	"ما هي الأعراض الأكثر شيوعاً والتي قد تكون بسبب ⁴ [[Disease]]؟"	1.
متوسط	[[Cause]]	"مريض يشتكي من ⁴ [[PatientComplaint]], وتظهر عليه علامات ⁴ [[Symptoms]], ما هو العامل المسبب الأكثر احتمالاً في هذه الحالة؟"	2.
متوسط	[[PhysicalExamination]]	"ما هي الفحوصات الجسدية التي يجب على الطبيب إجراؤها عند معاينة مريض يشتكي من ⁴ [[PatientComplaint]] وتظهر عليه علامات ⁴ [[Symptoms]]؟"	3.
صعب	[[Cause]]	"[[DemographicInformation]] تعاني من ⁴ [[Symptoms]]. ما هو التشخيص الأكثر احتمالاً في هذه الحالة؟"	4.

فمن أجل الحصول على سؤال سهل، يمكننا وضع خيارات خاطئة بعيدة عن الخيار الصحيح بحيث يمكن للطالب استبعاد الخيار الخاطئ بسهولة، ومن أجل الحصول على سؤال أكثر صعوبة، يتم اختيار خيارات خاطئة قريبة من الخيار الصحيح بحيث لا يمكن للطالب استبعاد الخيارات الخاطئة بسهولة.

5. التجربة (Experiment)

تم اجراء التجربة على كتاب (A-z common symptom answer guide (Wasson et al., 6, 2004), وهو مرجع طبي في مجال طب الأسرة⁷ يتميز بأنه ذو بنية شبه مهيكلة (semi-structured), حيث تتشابه فصول الكتاب مع بعضها من حيث الشكل الخارجي. تم تصميم قناع الفصل بناءً على البنية الشبه مهيكلة للمرجع (شكل 7).



شكل 7. شكل تمثيلي لـ "قناع الفصل" المستخدم في التجربة.

تم تمرير "قناع الفصل" على جميع فصول الوثيقة بهدف استخراج شبكات المعرفة من جميع فصول الوثيقة (شكل 8). تم تصميم قوالب الأسئلة باللغة العربية مع اسناد درجة صعوبة لكل قالب (جدول 3). ثم تم تمرير كل قالب على شبكة المعرفة من أجل إيجاد قيم المتحولات الموجودة في القالب والحصول على

⁶ بالنسبة لسياسة الخصوصية، يسمح المؤلف باستخدام الكتاب للأغراض الغير تجارية والشخصية (Wasson et al., 2004, p. 5).

⁷ يتألف الكتاب من 75 فصلاً تتحدث عن أشهر الشكاوى والأمراض المسببة لها، يبلغ عدد صفحات الكتاب 240 صفحة

تمت مقارنة أداء⁸ النظام مع أداء نموذج ChatGPT-4o (Plus⁹)، حيث قمنا بطلب توليد 114 سؤال (case-based) من المرجع العلمي، وذلك باستخدام العبارة التوجيهية (Prompt) التالية:

“Generate 114 Arabic case-based questions from this medical reference, distractors should be extracted from the same chapter.”

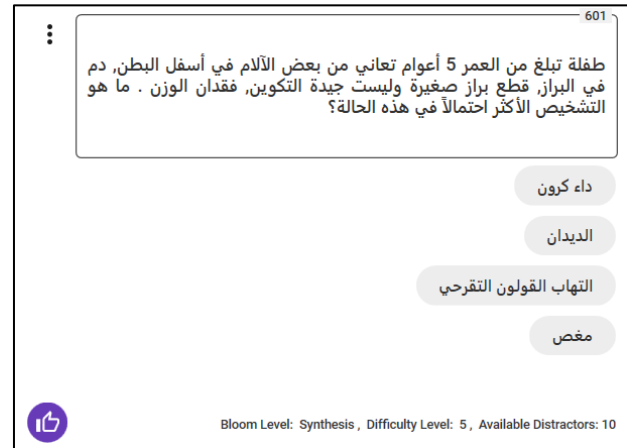
استطاع نموذج اللغة الكبير توليد حوالي 16 سؤال (case-based) في الدقيقة، أما بالنسبة للنظام المقترح، فقد تم ال114 سؤال بمعدل 14 سؤال في الثانية (تم احتساب زمن استخراج نموذج المعرفة في هذه المقارنة). تم عرض عينة عشوائية من الأسئلة على طبيبين مختصين وطلب منهم اسناد قيمة ضمن المجال [1-10] لكل سؤال مع الأخذ بعين الاعتبار كون السؤال صحيحاً وسليماً لغوياً، وبحسب تقييم الخبراء، فإن 80% من الأسئلة اعتبرت أسئلة صحيحة وذات درجات صعوبة مقبولة. تم قياس نسبة التوافق بين تقييمات الخبراء وفق نموذج Fleiss' kappa، فتم الحصول على نسبة 0.81، مما يعني أن اختيارات الخبراء كانت متوافقة، وأن الأسئلة الناتجة يمكن الاستفادة منها لأغراض التعليم والتقييم.

7. الخلاصة (Conclusion)

يهدف هذا العمل إلى مساعدة الطلاب العرب الغير ناطقين باللغة الإنكليزية عند دراستهم للمراجع العلمية التخصصية المكتوبة باللغة الإنكليزية. وذلك من خلال تطوير آلية توليد أسئلة باللغة العربية يمكن للطلاب الاستفادة منها لتقييم مدى فهمه للمحتوى العلمي. ويمكن للمدرسين الاستفادة من هذه

⁸ أما من حيث جودة الأسئلة، فقد تمت ملاحظة أن الأسئلة التي قام ChatGPT-4o (Plus) بتوليدها تعتبر أسئلة ذات جودة عالية (تقييم غير موضوعي - subjective evaluation).

⁹ يعتبر ChatGPT-4o بمثابة (state-of-the-art) بالنسبة لنماذج اللغات الكبيرة (LLM)، حيث يمكن لهذا ال(Bot) معالجة ملفات pdf واستخراج الأسئلة منها. وقد تم الاشتراك بنموذج (Plus) المدفوع لإجراء التقييم في هذه التجربة.



شكل 9. مثال عن السؤال المولد باستخدام القالب رقم (4).

6. النتائج والتقييم (Results and Evaluation)

بلغ عدد العُقد (nodes) الكلي لشبكة المعرفة الخاصة بالكتاب حوالي 3661 عقدة، بمتوسط 73.4 عقدة لكل فصل، وبلغ عدد الوصلات (edges) حوالي 5505 وصلة، بمتوسط 69.6 وصلة لكل فصل. تم استخراج نموذج المعرفة خلال 5.5 - 7.1 ثوان تقريباً، ويمكننا ملاحظة أن سرعة استخراج نموذج المعرفة تعني أنه يمكن إعادة بناء نموذج المعرفة بسرعة كبيرة في حالة صدور إصدارات جديدة من المرجع العلمي (إضافة فصول جديدة للمرجع أو تحديث معلومات قديمة في الفصول الموجودة). تم استخراج عينة (subset) من شبكة المعرفة بهدف تقييمها، تم اسناد قيمة (1) للوصلة الصحيحة واسناد قيمة (0) للوصلة الخاطئة، تم الحصول على قيمة (Recall) 95%، ويعود سبب ذلك إلى حصول أخطاء في استخراج البيانات من بواسطة "قناع الفصل"، وفي بعض الأحيان يكون ذلك بسبب أخطاء في ترجمة عُقد شبكة المعرفة إلى اللغة العربية.

بالنسبة لتوليد الأسئلة، فقد بلغ عدد الأسئلة المستخرجة من المرجع حوالي 3282 سؤال، بمتوسط 43.7 سؤال لكل فصل، وبزمن قدره 1.7 ثانية، وتم توليد 114 سؤال من الأسئلة ذات القصة السريرية (case-based).

- Science, Machine Learning and Artificial Intelligence*, 7–12.
<https://doi.org/10.1145/3484824.3484882>
- Alghamdi, E. A., Masoud, R. I., Alnuhait, D., Alo-mairi, A. Y., Ashraf, A., & Zaytoon, M. (2024). *AraTrust: An Evaluation of Trustworthiness for LLMs in Arabic* (arXiv:2403.09017). arXiv.
<https://doi.org/10.48550/arXiv.2403.09017>
- Alhashedi, S., Suaib, N. M., & Bakri, A. (2022). *Arabic Automatic Question Generation Using Transformer Model*. EasyChair.
https://easychair.org/publications/preprint_download/tzZ2
- Alhazmi, E., Sheng, Q. Z., Zhang, W. E., Zaib, M., & Alhazmi, A. (2024). *Distractor Generation for Multiple-Choice Questions: A Survey of Methods, Datasets, and Evaluation* (arXiv:2402.01512). arXiv.
<https://doi.org/10.48550/arXiv.2402.01512>
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain*. Longman New York.
- CH, D. R., & Saha, S. K. (2020). Automatic Multiple Choice Question Generation From Text: A Survey. *IEEE Transactions on Learning*
- الخدمة لتسريع عملية وضع الأسئلة الامتحانية. لم نتمكن من رصد أي عمل تناول مسألة توليد الأسئلة متعددة الخيارات باللغة العربية من المراجع العلمية التخصصية, تعتمد المنهجية المقترحة على كون كثير من المراجع العلمية تحتوي على بنية شبه مهيكلة (semi-structured), وبالتالي, تم اعتماد منهجية تعتمد على السمات البنوية للوثيقة (layout-based) بهدف استخراج نموذج معرفة غني يشمل جميع مواضيع المرجع العلمي. تم اعتماد منهجية (template-based) لاستخراج الأسئلة من نموذج المعرفة, تم اجراء تجربة على مرجع تخصصي في مجال طب الأسرة, بلغت دقة استخراج نموذج المعرفة 95%, وبلغ عدد الأسئلة الناتجة حوالي 3200 سؤال, منها 114 سؤال ذو قصة سريرية (case-based) بزمن كلي أقل من 10 ثوان. وبحسب تقييم الخبراء, فإن الأسئلة الناتجة تعتبر أسئلة جيدة وذات درجات صعوبة مقبولة يمكن استخدامها لأغراض التعليم والتقييم.
- ## 8. المراجع (References)
- Abdurahman, M., Darari, F., Lesmana, H., Hartopo, M., Rhesa, I., & Lumban Tobing, B. C. (2021). Lex2KG: Automatic Conversion of Legal Documents to Knowledge Graph. *2021 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, 1–5.
<https://doi.org/10.1109/ICACSIS53237.2021.9631310>
- Alazani, S. A., & Mahender, C. N. (2022). Rule Based Question Generation for Arabic Text: Question Answering System. *Proceedings of the International Conference on Data*

- Language Models in Arabic* (arXiv:2309.12053). arXiv. <https://doi.org/10.48550/arXiv.2309.12053>
- Jiang, B., Gu, M., & Du, Y. (2023). Recent Advances in Intelligent Textbooks for Better Learning. In H. Niemi, R. D. Pea, & Y. Lu (Eds.), *AI in Learning: Designing the Future* (pp. 247–261). Springer International Publishing. https://doi.org/10.1007/978-3-031-09687-7_15
- Khondaker, M. T. I., Waheed, A., Nagoudi, E. M. B., & Abdul-Mageed, M. (2023). GPTAraEval: A Comprehensive Evaluation of ChatGPT on Arabic NLP. In H. Bouamor, J. Pino, & K. Bali (Eds.), *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing* (pp. 220–247). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2023.emnlp-main.16>
- Kumar, S., Chauhan, A., & Kumar, P. (2022). Augmenting eBooks with with recommended questions using contrastive fine-tuned T5. In Md. S. Akhtar & T. Chakraborty (Eds.), *Proceedings of the 19th International Conference on Natural Technologies*, 13(1), 14–25. IEEE Transactions on Learning Technologies. <https://doi.org/10.1109/TLT.2018.2889100>
- CH, D. R., & Saha, S. K. (2023). Generation of Multiple-Choice Questions From Textbook Contents of School-Level Subjects. *IEEE Transactions on Learning Technologies*, 16(1), 40–52. IEEE Transactions on Learning Technologies. <https://doi.org/10.1109/TLT.2022.3224232>
- Dijkstra, R. (2022). *Quiz Generation using Generative Pre-trained Transformers*.
- Dresscher, L., Alpizar-Chacon, I., & Sosnovsky, S. (2021). Generation of assessment questions from textbooks enriched with knowledge models. *CEUR Workshop Proceedings*, 2895, 45–59. <https://ceur-ws.org/Vol-2895/paper10.pdf>
- Han, Z. (2022). *Unsupervised multilingual distractor generation for fill-in-the-blank questions*. <https://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-476855>
- Huang, H., Yu, F., Zhu, J., Sun, X., Cheng, H., Song, D., Chen, Z., Alharthi, A., An, B., He, J., Liu, Z., Zhang, Z., Chen, J., Li, J., Wang, B., Zhang, L., Sun, R., Wan, X., Li, H., & Xu, J. (2023). *AceGPT, Localizing Large*

- P. J. Muñoz-Merino, T. De Laet, A. Ortega-Arranz, & T. Farrell (Eds.), *Educating for a New Future: Making Sense of Technology-Enhanced Learning Adoption* (pp. 272–284). Springer International Publishing.
https://doi.org/10.1007/978-3-031-16290-9_20
- Olney, A. M. (2023). Generating multiple choice questions from a textbook: Llms match human performance on most metrics. *AIED Workshops*.
<https://ceur-ws.org/Vol-3487/paper7.pdf>
- Pan, L. (2022). *Towards Generating Deep Questions from Text* [PhD Thesis, National University of Singapore (Singapore)].
<https://search.proquest.com/openview/5fddb52b101fb5a0fed64706ea8f87/1?pq-origsite=gscholar&cbl=2026366&diss=y>
- Patil, S., Chavan, L., Mukane, J., Vora, D., & Chitre, V. (2022). State-of-the-Art Approach to e-Learning with Cutting Edge NLP Transformers: Implementing Text Summarization, Question and Distractor Generation, Question Answering. *International Journal of Advanced Computer Science and Applications*, 13(1).
Language Processing (ICON) (pp. 109–115). Association for Computational Linguistics.
<https://aclanthology.org/2022.icon-main.15>
- Kusuma, S. F., Siahaan, D. O., & Fatichah, C. (2022). Automatic question generation with various difficulty levels based on knowledge ontology using a query template. *Knowledge-Based Systems*, 249, 108906.
<https://doi.org/10.1016/j.knosys.2022.108906>
- Leo, J., Kurdi, G., Matentzoglou, N., Parsia, B., Sattler, U., Forge, S., Donato, G., & Dowling, W. (2019). Ontology-Based Generation of Medical, Multi-term MCQs. *International Journal of Artificial Intelligence in Education*, 29(2), 145–188.
<https://doi.org/10.1007/s40593-018-00172-w>
- Naous, T., Ryan, M. J., Ritter, A., & Xu, W. (2024). *Having Beer after Prayer? Measuring Cultural Bias in Large Language Models* (arXiv:2305.14456). arXiv.
<https://doi.org/10.48550/arXiv.2305.14456>
- Nguyen, H. A., Bhat, S., Moore, S., Bier, N., & Stamper, J. (2022). Towards Generalized Methods for Automatic Question Generation in Educational Domains. In I. Hilliger,

- ptom-Answer-Guide-ebook/dp/B00UYDF56M
- Willert, N., & Thiemann, J. (2024). Template-Based Generator for Single-Choice Questions. *Technology, Knowledge and Learning*, 29(1), 355–370. <https://doi.org/10.1007/s10758-023-09659-5>
- Pistol, I., Trandabăț, D., & Răschip, M. (2018). Medi-Test: Generating Tests from Medical Reference Texts. *Data*, 3(4), Article 4. <https://doi.org/10.3390/data3040070>
- Rahim, T. N. T. A., Aziz, Z. A., Rauf, R. H. A., & Shamsudin, N. (2017). Automated exam question generator using genetic algorithm. *2017 IEEE Conference on E-Learning, e-Management and e-Services (IC3e)*, 12–17. <https://doi.org/10.1109/IC3e.2017.8409231>
- Sengupta, N., Sahu, S. K., Jia, B., Katipomu, S., Li, H., Koto, F., Marshall, W., Gosal, G., Liu, C., Chen, Z., Afzal, O. M., Kamboj, S., Pandit, O., Pal, R., Pradhan, L., Mujahid, Z. M., Baali, M., Han, X., Bsharat, S. M., ... Xing, E. (2023). *Jais and Jais-chat: Arabic-Centric Foundation and Instruction-Tuned Open Generative Large Language Models* (arXiv:2308.16149). arXiv. <https://doi.org/10.48550/arXiv.2308.16149>
- Wasson, J., Walsh, B. T., LaBrecque, M. C., Sox, H. C., Pantell, R., & Oransky, I. (2004). *A-Z Common Symptom Answer Guide* (1st ed.). McGraw-Hil. <https://www.amazon.com/Z-Common-Sym>