

## بناء بنك أسئلة طبية متعددة الخيارات باستخدام منهجية دلالية لتوليد الخيارات الخاطئة مع قياس درجة الصعوبة

### الملخص

يقدم البحث حلاً للصعوبات الشائعة التي تواجه أنظمة توليد الأسئلة متعددة الخيارات آلياً في المجال الطبي وذلك من خلال توليد عدد أكبر من الإجابات الخاطئة الجيدة (Distractors) مما يؤدي لزيادة كبيرة في عدد الأسئلة المولدة مع إتاحة إمكانية حذف الأسئلة البسيطة وإمكانية تجميع واختيار الأسئلة بحسب الموضوع الطبي. وقد تم ذلك من خلال 1. استخدام شبكة العلاقات الدلالية التي تربط بين المفاهيم الطبية في نظام اللغة الطبية الموحد (UMLS) كمصدر مساعد جنباً إلى جنب مع شبكة العلاقات الهرمية (العائلية) التقليدية مما سمح بتوليد عدد كبير من الخيارات الخاطئة ذات صلة قرابة عالية بالخيار الصحيح 2. تطوير خوارزمية تقوم بإسناد درجة صعوبة للأسئلة المتولدة وبالتالي تمكّننا من استئصال الأسئلة شديدة السهولة. حققت المنهجية المستخدمة زيادة كبيرة في عدد الأسئلة بنسبة 100% عن عدد الأسئلة المتولدة بالطرق التقليدية وتمّ بناء بنك أسئلة مكون من 1,000,000 سؤالاً طبياً متعدد الخيارات مستخرجة من 34 مصدر للبيانات الطبية (أنطولوجيات، قواعد معرفة، قواميس)، وتمّ إسناد درجة صعوبة لكل سؤال مما يتيح ترتيبها واختيارها حسب درجة الصعوبة المرغوبة مع رسم كل سؤال بصنف دلالي مطابق للصنف الدلالي الخاص بالجواب الصحيح مما يتيح إمكانية تجميع واختيار الأسئلة حسب الصنف الدلالي (مرض، عرض، إنزيم، دواء، الخ..).

**كلمات مفتاحية:** الأسئلة متعددة الخيارات، قاعدة معرفة، أنطولوجي، معجم، نظام اللغة الطبية الموحد UMLS، التعليم، الطب.

## **Building Multiple Choice Medical Question Bank Using Semantic Approach to Generate More Distractors Along with Question Difficulty Measurement**

---

### **Abstract**

This work solves common challenges that face the traditional Automatic Question Generation Systems in medical domain. By finding a new way to generate more good distractors, the rate of the overall generated questions has been increased, and by measuring the difficulty for each question, the easy questions have been removed. We also have tagged every question with the correct-answer's semantic type so it became possible to group the generated questions and query them according to their medical semantic types. This was done by: 1. Using medical semantic network within the Unified Medical Language System (UMLS) as a second resource for generating good distractors, along with the traditional hierarchy network that connects each concept with its parents and children, so adding this new resource allows for generating more good distractors and therefore generating more good questions, 2. Developing an algorithm to measure the difficulty of each question so we can suppress the easy questions. This approach enhances the quality of the generated questions and scales up the quantity by 100% compared to traditional approaches. We have extracted 1,000,000 multiple choice questions in medical domain from 34 medical resources (Ontologies, knowledge bases, thesauruses), with each question having a difficulty score so the examiner can sort, remove and pick the questions according to the desired difficulty level, furthermore, each question is tagged with one or more semantic type, so it is possible to group the questions according to their medical semantic type (Disease, Symptom, Enzyme, Drug, etc...).

**Keywords:** Multiple Choice Questions, Knowledge base, Ontology, Thesaurus, UMLS, Education, Medicine.

---

## مقدمة

تتميز الأسئلة متعددة الخيارات عن الأسئلة مفتوحة الإجابة (Free response) بأنها سهلة التصحيح ولكنها تتطلب جهداً إضافياً من قبل المصمم من أجل اختيار الإجابات الخاطئة المناسبة بحيث لا يمكن للطلاب ضعيف التحضير أن يستبعد الخيارات الخاطئة بسهولة ويزيد من فرصة معرفة الجواب الصحيح من خلال الاختيار العشوائي. ويعتبر توليد الخيارات الخاطئة (Distractors Generation) من أصعب التحديات التي تواجه أنظمة استخراج الأسئلة متعددة الخيارات [3]، وتركز بعض الدراسات على هذا الموضوع فقط [4]، وكذلك تعتبر بنوك الأسئلة الصغيرة نقطة ضعف بالنسبة للعملية التعليمية حيث أنها مع مرور الوقت تصبح مكررة وبالتالي تصبح غير قادرة على تمييز وتقييم الطلاب بكفاءة، وكذلك فإن تجديدها باستمرار يعتبر أمراً مكلفاً. يمكن لأنظمة توليد الأسئلة آلياً أن تدعم العملية التعليمية من عدة نواحي: 1. تسمح للمدرس أن يضيف الأسئلة المولدة آلياً إلى بنك الأسئلة الخاص بالمقرر التعليمي. 2. يمكن للطلاب الاستفادة من إتاحة عدد كبير من الأسئلة المولدة آلياً وتقييم أدائهم الامتحاني من خلال مقارنة إجاباتهم مع الإجابة الصحيحة. 3. استخراج أسئلة تغطي المقرر الدراسي (الجديد أو المعدل) بشكل كامل وبكلفة منخفضة جداً مقارنة مع الطرق التقليدية. 4. يمكن لنظام توليد الأسئلة آلياً أن يقوم باستخراج أسئلة تتناسب مع أهداف المقرر التعليمية.. ففي المجال الطبي مثلاً يُطلب استخراج أسئلة تختبر قدرة الطالب على تشخيص حالة سريرية بالنظر إلى الأعراض والتاريخ المرضي، أو يمكن مثلاً اختبار قدرة الطالب على اقتراح الدواء المناسب للحالة المرضية مع الأخذ بعين الاعتبار كون المريض يتناول أدوية معينة وبالتالي يجب على الطالب أن يراعي احتمالية وجود تداخلات دوائية واستبعاد بعض الأدوية. ولهذا الغرض يجب أن يتم التنسيق بين مصمم النظام والخبير المختص والافتقار على نمط الأسئلة المرغوب توليدها آلياً كما في (شكل 2). يتم استخدام الأنطولوجيات (Ontologies) بهدف استخراج الخيارات الخاطئة للسؤال في مجال معين، تحتوي الأنطولوجي بشكل أساسي على: المفاهيم وتعريفها (إن وجدت)، والروابط من نوع is\_a (أب-ابن) التي تربط بين المفاهيم. ومن مساوئ هذه الطريقة أنها لا يمكنها أن تولد سؤالاً جيداً إذا لم تجد العدد الكافي من الأجوبة الخاطئة ضمن شبكة العلاقات الهرمية في الأنطولوجي (Ontology)، لذلك نعتقد أن الطرق الحالية تخضع لقيود تحد من كمية ونوعية الأسئلة المتولدة آلياً وفي هذا البحث نقدم مقارنة تهدف إلى تخفيف هذه القيود ورفع عدد

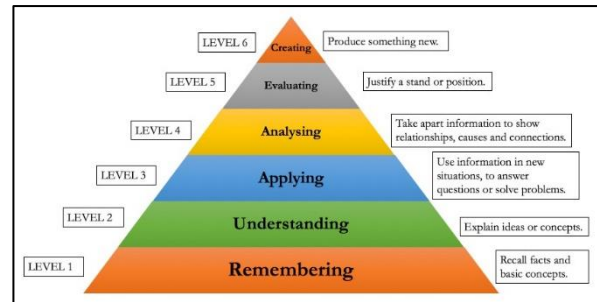
تعتبر عملية وضع الأسئلة متنوعة الصعوبة والتي تختبر فهم الطالب للمادة العلمية عملية مكلفة من حيث الوقت والجهد والمال [1]، وبالنظر إلى العدد الكبير والمتزايد للطلاب ومع تطور أساليب التعلم عن بعد (مثل<sup>1</sup> MOOC)، فقد أصبحت هناك حاجة كبيرة إلى وضع الكثير من الأسئلة وإنشاء بنوك الأسئلة التي تغطي المقررات الدراسية بشكل كامل، ويتطلب وضع الأسئلة التعليمية الجيدة من قبل المدرس مستوى عالياً من الإدراك والتركيز [2]. وقد بدأ العمل على استخراج الأسئلة آلياً من مختلف أنواع المصادر النصية [2,3]:

1. المهيكلة (Structured) مثل الأنطولوجيات

(Ontologies)

2. غير المهيكلة (Unstructured)، مثل النصوص (Text).

يتبع السؤال المولد آلياً أحد الأنواع التالية: 1. أسئلة تعتمد على الاستدكار (عرف، صف، أذكر مثلاً..). 2. أسئلة مفتوحة الإجابة (ما، أين، لماذا، متى، من). 3. أسئلة الاختيار من متعدد [1]. ويعتبر هرم بلوم (شكل 1) مرجعاً لتصنيف الأهداف التعليمية وأنواع الأسئلة، وهو مقسم إلى 6 مستويات تتدرج من الأبسط إلى الأكثر تعقيداً كما يلي: 1. التذكر (عرف، عدد). 2. الفهم (اشرح، لخص، قارن، صف). 3. التطبيق (استخدم، نفذ، أنجز، استعرض). 4. التحليل (استكشف، اختر، حل). 5. التقييم (ماذا نتوقع/برأيك، كيف تحكم/تري). 6. التأليف والاستنتاج (صمم، طور، أنشئ، ادمج).



شكل 1. هرمية بلوم Bloom's Taxonomy<sup>2</sup>

يهدف البحث في هذا المجال إلى تطوير نظام استخراج أسئلة بشكل آلي تستهدف مستويات عليا من هرمية بلوم [5].

<sup>2</sup> مصدر الصورة - <https://technologyforlearners.com/applying-blooms-taxonomy-to-the-classroom>

<sup>1</sup> Massive Open Online Courses

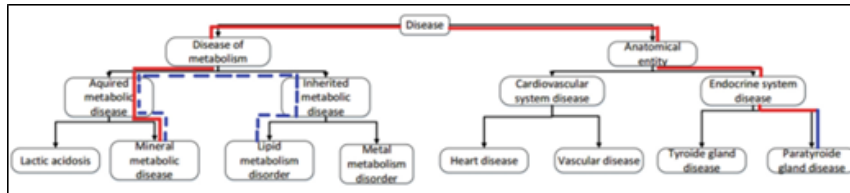
**Template 1: What is the most likely diagnosis?**

A Patient-demographic patient with {History}<sup>\*a</sup> presents with {Symptom}<sup>\*</sup>.  
What is the most likely diagnosis?

- A. Correct Disease  
B. Incorrect Disease  
C. Incorrect Disease

<sup>a</sup>\* = one or more entities.

شكل 2. سؤال اختيار من متعدد في المجال الطبي يتبع لقالب معين، مأخوذ من [8]



شكل 3. كيفية اختيار الجواب الخاطئ من الأنطولوجي. مأخوذ من [11]

الأنطولوجيات، فإذا لم يكن هناك عدد كافٍ من المفاهيم ذات صلة قرابة عالية بينها وبين المفهوم الهدف (الجواب الصحيح) فهذا يعني أننا لن نتمكن من انتاج سؤال جيد حول المفهوم الهدف أو أن عدد الأسئلة الممكن طرحها حول هذا المفهوم سيكون صغيراً نسبياً. في [8] تم استخدام قاعدة المعرفة EMMET<sup>5</sup> التي تربط بين المفاهيم الطبية بعلاقات من نوع (boarder/narrower) وبالتالي تم تحويلها إلى أنطولوجي ضخمة واستخراج الأسئلة الطبية متعددة الخيارات منها.

جميع الطرق التقليدية تمر بنفس المراحل لاقتراح الخيارات الخاطئة وهي كما يلي: يتم النظر إلى الجواب الصحيح (key) والبحث في الأنطولوجي:

1. إذا وجدت له 3 اخوة فبالتالي يمكن استخدامها كأجوبة خاطئة (Distractors). ويعتبر السؤال صعباً نسبياً لأن درجة القرابة بين الجواب الصحيح والأجوبة الخاطئة تعتبر أقرب ما يمكن.
2. إذا لم يكن عدد الأخوة كافٍ فيمكننا اقتراح أبناء العمومة كأجوبة خاطئة (وهكذا يصبح السؤال أقل صعوبة)
3. وهكذا بشكل تراجمي حتى يصبح لدينا 3 خيارات خاطئة.. وكلما ازدادت المسافة بين الجواب الصحيح والجواب الخاطئ فسوف تنقص درجة صعوبة السؤال.

أي أن المشترك بين الأبحاث السابقة هو استخدامها للأنطولوجيات واعتمادها فقط على علاقات الأب-ابن (is\_a) لاختيار الخيارات

الأسئلة المتولدة وزيادة جودتها. نذكر فيما يلي طريقة عمل الطرق التقليدية ونقاط ضعفها، نصف بعد ذلك المنهجية المقترحة للحل وكيفية إنجازها، ثم نستعرض النتائج ونناقشها، وفي النهاية نوضح بعض الجوانب التي توجد حاجة للعمل على تطويرها.

## دراسة مرجعية

يعتبر البحث في توليد الأسئلة آلياً مجالاً نشطاً [4]، ففي [3] تم عمل دراسة استقصائية (Survey) ذكر فيها أنه تم نشر 3422 ورقة مؤتمر، 2222 ورقة مجلة علمية، و2213 ورقة ورشات عمل، وأنه يوجد ما يقارب من 358 باحثاً يعملون في 13 مجموعة عمل في مجال توليد الأسئلة آلياً في مختلف المجالات ومختلف اللغات. وتعتبر الأسئلة متعددة الخيارات أسئلة نوعية ومهمة في المجال التعليمي، ففي المجال الطبي تم في [6] و [3] استخراج الأسئلة متعددة الخيارات آلياً بهدف تقديمها للطلاب في امتحان الأدوية، وذلك باستخدام قوالب معدة مسبقاً للأسئلة بالإضافة إلى أنطولوجي (Ontology) طبية، وفي [7] تم استخدام القوالب والأنطولوجي كذلك لتوليد الأسئلة متعددة الخيارات من قاعدة بيانات MeSH<sup>3</sup>. وتعتبر أنطولوجي HPO (Human Phenotype Ontology) هي الأكبر حجماً في المجال الطبي حيث تحوي على 18,000 مصطلح طبي تتحدث عن الأمراض<sup>4</sup> عند الإنسان مع تصنيفها، وترتبط المفاهيم فيها بعلاقات أب-ابن (is\_a) فقط، وهنا يظهر لنا قيد يحد من حجم بنك الأسئلة الممكن توليده من

<sup>5</sup> بيانات خاصة وغير متاحة.

<sup>3</sup> Medical Subject Headings

<sup>4</sup> <https://hpo.jax.org/app>

3. عدد كبير من العلاقات الدلالية والعلاقات العائلية (الهرمية) بين المفاهيم (12,000,000 علاقة)
4. كل مفهوم ينتمي إلى صنف دلالي (مثل: مرض، دواء، جين، حمض أميني، جهاز مخبري، الخ...) وبالتالي يمكننا رسم السؤال بنفس الصنف الدلالي للخيار الصحيح، وهذا يمكننا من تجميع الأسئلة حسب النوع الدلالي.
5. يمكننا اجتزاء هذه البيانات (subsetting) وفلترتها حسب الصنف الدلالي.

يمكننا استخدام شبكة العلاقات الدلالية وشبكة العلاقات الهرمية معاً لتوليد الخيارات الخاطئة والحصول على عدد كبير جداً من الأسئلة الطبية المنمطة دلاليًا حيث أن استخدام شبكة العلاقات الدلالية سيزيد من عدد الخيارات الخاطئة المتاحة لكل سؤال. ومن خلال رسم السؤال بالصنف الدلالي الخاص بالإجابة الصحيحة يصبح بإمكاننا تجميع الأسئلة حسب الموضوع (مثلاً: مجموعة الأسئلة الخاصة بالأمراض، مجموعة الأسئلة الخاصة بالأدوية، إلخ...). يتم أثناء توليد الأسئلة منح كل سؤال درجة صعوبة وذلك بحسب قوة ارتباط الخيارات الخاطئة بالخيار الصحيح، باختصار، تزداد صعوبة السؤال كلما قل عدد الوصلات بين الخيار الصحيح والخيارات الخاطئة سواء في شبكة العلاقات الدلالية أو شبكة العلاقات العائلية.

يتم تمرير الأسئلة الناتجة إلى مرحلة معالجة أخيرة (Post-Processing) يتم فيها تعديل القيمة العددية الخاصة بدرجة صعوبة السؤال وذلك من خلال مقارنة التقارب النصي بين جسم السؤال والخيار الصحيح، وبين جسم السؤال والخيارات الخاطئة، وبين الخيار الصحيح والخيارات الخاطئة، وفق قواعد محددة. في النهاية، يصبح بالإمكان فلتره الأسئلة وفق معيارين: درجة صعوبة السؤال (مثلاً: نريد الأسئلة ذات درجة صعوبة أكثر من 70%)، والصنف الدلالي للسؤال (مثلاً: نريد الأسئلة المتعلقة بأمراض العين).

### التجربة

تمر التجربة بالمراحل التالية:

الخاطئة، ونعتقد أن إضافة مصادر جديدة لتوليد الخيارات الخاطئة سيحسن من نوعية وكمية الأسئلة المتولدة آلياً، وسيساهم في تخفيف أثر القيود التي يفرضها استخدام علاقات is\_a فقط لتوليد الخيارات الخاطئة في أسئلة الاختيار من متعدد.

### المنهجية المقترحة

يمكن زيادة عدد الأسئلة كلما ازدادت الخيارات الخاطئة، فمن أجل كل ثنائية (سؤال-جواب صحيح)، نحتاج إلى 3 خيارات خاطئة لتوليد سؤال واحد، وفي حال وجدت 4 خيارات خاطئة فهذا يعني إمكانية توليد 4 أسئلة مختلفة، ومن أجل 5 خيارات خاطئة يمكننا توليد 10 أسئلة، وهكذا.. وفق العلاقة:

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

حيث  $n \geq 3$  هو عدد الأجوبة الخاطئة المتاحة للسؤال Q ذو الإجابة A.

تعد شبكة العلاقات الدلالية التي تربط بين المفاهيم المختلفة بعلاقات دلالية مصدراً جيداً لتوليد الخيارات الخاطئة، لنأخذ علاقة (متعلق ب) كمثال، فإذا كان الجواب الصحيح هو (المرض-أ)، ولديه علاقة (متعلق ب) تربط بينه وبين (المرض-ب)، فإمكاننا استخدام (المرض-ب) كخيار خاطئ (Distractor) للسؤال الذي نحن بصدد توليده، وهكذا بالنسبة لبقية الخيارات. وللحصول على بنك من الأسئلة متعدّدة الخيارات ذات جودة مقبولة يتم تمرير الأسئلة عادةً على مصنف (Filter) يقوم بإسناد درجة صعوبة لكل سؤال وبالتالي إتاحة إمكانية اختيار الأسئلة حسب درجة صعوبتها أو حذف الأسئلة السهلة نسبياً واختيار الأسئلة الصعبة ومتوسطة الصعوبة لإدراجها ضمن بنوك الأسئلة الخاصة بالمقرر الدراسي.

تهدف المنهجية المقترحة إلى توليد عدد كبير من الأسئلة الطبية من مواضيع مختلفة مع قياس صعوبتها بحيث يسهل على الخبير اختيار عينة منها لاستخدامها في المادة الامتحانية. تم اختيار مجموعة البيانات Corpora المتاحة في "نظام اللغة الطبية الموحد" (UMLS) [9]، والتي تعتبر مصدراً نوعياً وغنياً بالبيانات الطبية، فهي تحتوي على:

1. عدد كبير من مصادر البيانات (60 مصدر)
2. عدد كبير جداً من المفاهيم الطبية (900,000 مفهوم)

<sup>6</sup> مثل العلاقة disease\_has\_associated\_disease في الشبكة الدلالية في نظام UMLS

<sup>7</sup> Unified Medical Language System  
<https://www.nlm.nih.gov/research/umls/licensedcontent/umlsknowledge/sources.html>

صف، تم اختيار 170 صنف منها كصنف مرغوب<sup>10</sup>، وتم إقصاء المفاهيم التي لا تنتمي إلى الصفوف المرغوبة (مثل (Food).

2. اختيار المفاهيم المفضلة (Preferred) من جدول mrconso وجلب تعاريفها من جدول mrdef وجلب صنفها الدلالي من جدول mrsty ودمجها معاً للحصول على جدول بيانات جديد يحتوي بشكل رئيسي على: (المفهوم، التعريف، الصف الدلالي).

وبالتالي يصبح عدد التسجيلات (Records) في الجدول الجديد أكثر من 180,000 تسجيلية.

3. استخراج ثنائيات الأسئلة والخيارات الصحيحة (سؤال، خيار صحيح) مع وسمها بالصنف الدلالي الخاص بالجواب الصحيح من خلال مقابلة ما يلي: مقابلة المفهوم (Concept) بالجواب الصحيح، مقابلة تعريف المفهوم (Definition) بجسم السؤال، إسناد الصف الدلالي للمفهوم (Semantic type) إلى الصف الدلالي للسؤال، فعلى سبيل المثال: إذا كان المفهوم ينتمي إلى صف إنزيم (Enzyme)، فإننا نعتبر السؤال عن الإنزيمات.

4. استخراج الأجوبة الخاطئة (Distractors) وإسناد درجة صعوبة لكل سؤال، وذلك من خلال جلب المفاهيم الأخوة (Siblings) من شبكة علاقات القرابة (أب-ابن) (من جدول mrhier)، وجلب المفاهيم المرتبطة دلاليًا من شبكة العلاقات الدلالية (من جدول mrrel)، بحيث يبلغ عدد المفاهيم المستخرجة على الأقل 3 مفاهيم (Concepts) كحد أدنى، فمثلاً، إذا وجد للمفهوم أخ واحد فقط في شبكة العلاقات الهرمية التقليدية، فإننا نبدأ البحث في شبكة العلاقات الدلالية بغية الحصول على المفاهيم المرتبطة المتبقية (نحتاج في هذه الحالة إلى مفهومين إضافيين على الأقل

1. اجتزاء مجموعة البيانات المرغوبة من "نظام اللغة الطبية الموحد" (UMLS) وتحميلها في قاعدة معطيات MySQL: تم اختيار 34 مجموعة بيانات (Dataset)<sup>8</sup> للعمل عليها بهدف الحصول على أكبر عدد ممكن من الأسئلة. نجد في (جدول 1) أهم مصادر البيانات المختارة مع عدد المفاهيم الطبية الواردة ضمنها.

عدد المفاهيم الطبية الموجودة ضمنه	اسم المصدر
1,983,467	NCBI Taxonomy
454,624	Medical Subject Headings
416,534	US Edition Of SNOMED CT
241,639	UMLS Metathesaurus
169,826	NCI Thesaurus
71,255	Gene Ontology
17,913	Human Phenotype Ontology
10,055	Drug Bank

جدول 1. أهم مصادر البيانات المستخدمة في البحث

عند انتهاء عملية الاجتزاء (Subsetting) نقوم بإنشاء قاعدة البيانات MySQL الخاصة بالتجربة ونقوم بتهجير محتويات المجموعة المجتزأة (Subset) إلى جداول MySQL<sup>9</sup>.

2. دمج جداول البيانات للحصول على جداول جديدة لاستخدامها في توليد الأسئلة: يبلغ عدد المفاهيم المستخرجة أكثر من 4,000,000 مفهوم (Concept) من دون تكرار، وسنقوم بتقليص المجموعة من خلال:

1. اختيار المفاهيم التي تنتمي إلى صف دلالي مرغوب، حيث يبلغ عدد الصفوف الدلالية في بيانات UMLS حوالي 204

<sup>9</sup> يمكن الاطلاع على تقرير استخراج البيانات من هنا

<https://drive.google.com/file/d/1wd5rhQnQ7zzNjFaskFqiQHvOnTXZDm5R/view?usp=sharing>

<sup>10</sup> يمكن الاطلاع عليها من خلال العنوان

[https://drive.google.com/file/d/19Of9TXVccv7J\\_UrEVp4OGRXmQqlCkh/view?usp=sharing](https://drive.google.com/file/d/19Of9TXVccv7J_UrEVp4OGRXmQqlCkh/view?usp=sharing)

<sup>8</sup> يمكن إعادة اشتقاق هذه المجموعة بشكل مباشر من خلال ملف الإعدادات الخاص بهذه التجربة والموجود في العنوان

[https://drive.google.com/file/d/1Sgy\\_anKeLrgS\\_-pmCRB0p-s0B1eZ1nOO/view?usp=sharing](https://drive.google.com/file/d/1Sgy_anKeLrgS_-pmCRB0p-s0B1eZ1nOO/view?usp=sharing)

An anomaly of the hair follicles of the skin that typically presents as small, rough, brown folliculocentric papules distributed over characteristic areas of the skin, particularly the outer-upper arms and thighs. [PMID:22628989, PMID:27194977, PMID:3004...]

A: Cutaneous Involvement  
B: Epithelial hyperplasia (disorder)  
C: Keratosis pilaris  
D: Hyperkeratosis (disorder)

correct answer: Keratosis pilaris

question type: Disease or Syndrome

degree of difficulty: 6

شكل 4. مثال عن السؤال المتوكد

5. معالجة نهائية (Post-Processing): نقوم بتعديل درجات الصعوبة من خلال مقارنات داخلية بين نصوص كل من (جسم السؤال، الخيار الصحيح، الخيارات الخاطئة) وفق ما يلي:
1. يتم إنقاص 3 درجات إذا ورد الجواب الصحيح (بشكل جزئي أو كامل) في نص السؤال ولم يرد أي جواب خاطئ في نص السؤال، (يمكن التخفيف من هذا الأثر من خلال استبدال الجواب الصحيح بضمير it حال وروده بشكل كامل في نص السؤال).
  2. تتم زيادة درجة واحدة في حال ورد أحد الأجوبة الخاطئة في نص السؤال ولم يرد الجواب الصحيح فيه.
  3. تتم زيادة درجتين إذا ورد أكثر من جواب خاطئ في نص السؤال ولم يرد الجواب الصحيح فيه.
  4. لا تتغير الدرجة إذا لم يرد الجواب الصحيح أو أي جواب خاطئ في نص السؤال.
- لتكوين السؤال)، والأولية للمفاهيم التي تنتمي إلى نفس الصنف الدلالي، فإذا لم يبلغ عدد المفاهيم 3، نبدأ بالانتقال إلى المفاهيم التي يمكن الانتقال إليها بخطوة إضافية من المفاهيم التي تم اختيارها، ولا نبتعد في البحث أكثر من ذلك، لأننا نفترض بأنه إذا زادت المسافة بين الجواب الصحيح والجواب الخاطئ عن وصلتين فإن درجة صعوبة السؤال ستصبح منخفضة جداً. وإذا لم نحصل على 3 مفاهيم نقوم بوسم التسجيل (record) في قاعدة البيانات على أنها غير صالحة لاستخراج سؤال متعدد الخيارات متعلق بها.
- نقوم من أجل كل سؤال ناتج بإسناد قيمة عددية (score) تساوي مجموع عدد الوصلات التي تفصل بين الجواب الصحيح وكل جواب خاطئ، وبالتالي فإننا نعرف درجة صعوبة السؤال بشكل مبدئي على أنها متعم القيمة (score) إلى 10، أي (10 - score)، وذلك لأن العلاقة بين قيمة (score) ودرجة صعوبة السؤال هي علاقة عكسية، فكلما ازداد طول الوصلات بين الإجابة الصحيحة والأجوبة الخاطئة (أي، قيمة score) فسوف تنخفض درجة صعوبة السؤال، وعند الانتقال إلى المرحلة النهائية يتم تعديل هذه القيمة وفق قواعد اختبار مطورة بهدف الحصول على قيمة نهائية لدرجة صعوبة السؤال. وعند الانتهاء من هذه المرحلة يصبح لدينا 1,000,000 سؤال مستخرج من جدول التعاريف المذكور في الخطوة 2. (شكل 4) يبين مثالاً عن السؤال المتوكد.

### النتائج والمناقشة

تم استخدام "نظام اللغة الطبية الموحد" (UMLS) للاستفادة من حجم البيانات الطبية الكبير فيه وللاستفادة من شبكة العلاقات الدلالية بهدف توسيع دائرة البحث عن الأجوبة الخاطئة (Distractors). تم الحصول على 1,000,000 سؤال منمطاً دلاليّاً، يمكن ترتيب الأسئلة حسب درجة صعوبتها (مقياس من 1 إلى 10) ويمكن تجميع الأسئلة حسب صنفها الدلالي (مرض، عرض، دواء، جين، إنزيم، الخ...) واستبعاد الأسئلة السهلة. تتفوق هذه الطريقة على الطريقة التقليدية من حيث عدد

## خلاصة

تم تطوير آلية اقتراح خيارات خاطئة (Distractors) للأسئلة متعددة الخيارات بالاعتماد على شبكة العلاقات الدلالية التي تربط بين المفاهيم الطبية الموجودة في قاعدة بيانات نظام اللغة الطبية الموحد (UMLS) وليس فقط على شبكة العلاقات الهرمية (أب-ابن) المستخدمة في الأبحاث المشابهة. وأظهر الاعتماد على العلاقات الدلالية زيادة كبيرة في عدد الأسئلة المستخرجة بسبب اقتراحها عدداً أكبر من الأجوبة الخاطئة (Distractors) التي يمكن استعمالها لتوليد عدد أكبر من الإجابات الخاطئة مع وسم الأسئلة دلاليًا، تم أيضاً قياس مدى صعوبة الأسئلة المتولدة مما يتيح إمكانية فلترتها واستبعاد الأسئلة السهلة. نحتاج إلى معالجة جسم السؤال بهدف إلغاء الارتباط بينه وبين الجواب الصحيح بهدف رفع درجة الصعوبة، ونهدف إلى تطوير النظام وجعله نظام اقتراح (Recommendation system) يقوم بطرح أسئلة تتناسب مع النص المدخل وتتناسب أيضاً مع الأهداف التعليمية.

## المراجع

- [1] S. Soni, P. Kumar and A. Saha, "Automatic Question Generation: A Systematic Review," *SSRN*, 2019.
- [2] L. Pan, W. Lei, T. S. Chua and M. Y. Kan, "Recent Advances in Neural Question Generation," *arXiv:1905.08949*, 2019.
- [3] J. Leo, G. Kurdi, N. Matentzoglou, B. Parsia, U. Sattler, S. Forge and W. Dowling, "Ontology-based generation of medical, multi-term MCQs," *International Journal of Artificial Intelligence in Education*, pp. 145-188, 2019.
- [4] T. Desai, P. Dakle and D. Moldovan, "Generating Questions for Reading Comprehension using

ونوعية الأسئلة الناتجة وذلك بسبب إضافة مصدر جديد للبحث عن الخيارات الخاطئة.

تعتبر الطريقة المتبعة في [10] رائدة الطرق (State of the art) فيما يتعلق باستخراج الأسئلة متعددة الخيارات من الأنطولوجيات بشكل مباشر، وتعتبر أنطولوجي Human Phenotype Ontology (HPO) الأكبر حجماً في المجال الطبي، إذ تحتوي على حوالي 18,000 مفهوم طبي تنتمي إلى عدد أصناف محدود (أمراض، جينات، الخ..). مرتبطة فيما بينها بعلاقات (أب-ابن) فقط. وتتميز الطريقة المقترحة<sup>11</sup> من حيث:

1. توسيع دائرة البحث عن الخيارات الخاطئة: في [10] تم الاقتصار على شبكة العلاقات العائلية (أب-ابن) لتوليد الخيارات الخاطئة، في حين تم الاعتماد في هذا البحث على شبكة العلاقات الدلالية بشكل رئيسي، جنباً إلى جنب مع شبكة العلاقات العائلية، فتم تحقيق زيادة في عدد الأسئلة أكثر من 100%.
2. تنميط السؤال دلاليًا وإسناد درجة صعوبة له مما يتيح إمكانية تجميع الأسئلة حسب الموضوع وترتيب الأسئلة حسب درجة الصعوبة واستبعاد الأسئلة السهلة.

## أعمال مستقبلية

تم توليد بنك ضخم من الأسئلة الطبية متعددة الخيارات بشكل آلي، عدد قليل من الأسئلة المستخرجة ينتمي إلى درجات صعوبة عالية، وذلك بسبب حصول تقاطع جزئي بين الإجابة الصحيحة وجسم السؤال في عدد كبير من الأسئلة، نحتاج إلى تطوير آلية تعالج جسم السؤال وتلغي الارتباط الظاهري بينه وبين الجواب الصحيح، وبالتالي يمكن رفع درجة صعوبة هذه الأسئلة. نهدف أيضاً إلى جعل النظام أكثر ديناميكية حتى نتمكن من اقتراح أسئلة تتناسب مع المحتوى التعليمي (المقرر الجامعي)، وأكثر من ذلك، نهدف إلى طرح أسئلة تتناسب مع أهداف المقرر الدراسي وذلك بالاشتراك مع المدرّس المختص، فمن أجل نص طبي ما، يهدف المقرر إلى تدريب الطالب على تشخيص الحالة المرضية، أو يطلب منه اقتراح الدواء المناسب (مع/دون مراعاة احتمالية وجود تداخلات دوائية بين الدواء المقترح وأحد الأدوية التي يتناولها المريض).

<sup>11</sup> علماً أن HPO موجودة ضمن قاعدة بيانات UMLS وتم استخدامها ضمناً في هذا البحث (جدول 1)



- Coherence Relations," in *the 5th Workshop on Natural Language Processing Techniques for Educational Applications*, 2018.
- [5] D. R. Ch and S. K. Saha, "Automatic multiple choice question generation from text: A survey," *IEEE Transactions on Learning Technologies*, 2018.
- [6] W. Wang, T. Hao and W. Liu, "Automatic question generation for learning evaluation in medicine," in *International conference on web-based learning*, Berlin, Heidelberg, 2007.
- [7] M. A. Lopetegui, B. A. Lara, P. Y. Yen, Ü. V. Çatalyürek and P. R. Payne, "A novel multiple choice question generation strategy: alternative uses for controlled vocabulary thesauri in biomedical-sciences education," in *AMIA Annual Symposium*, 2015.
- [8] G. R. Kurdi, Generation and mining of medical, case-based multiple choice questions, United Kingdom: The University of Manchester , 2020.
- [9] B. O, "The Unified Medical Language System (UMLS)," *integrating biomedical terminology*., 2004.
- [10] T. Alsubait, B. Parsia and U. Sattler, Ontology-based multiple choice question generation, KI-Künstliche Intelligenz, 2016.
- [11] M. Radovic, M. Tosic, D. Milosevic and D. Jankovic, "OntoCIP-an ontology of comprehensive integrative puzzle assessment method suitable for automatic question generation," in *International Conference on Interactive Collaborative Learning*, 2017.