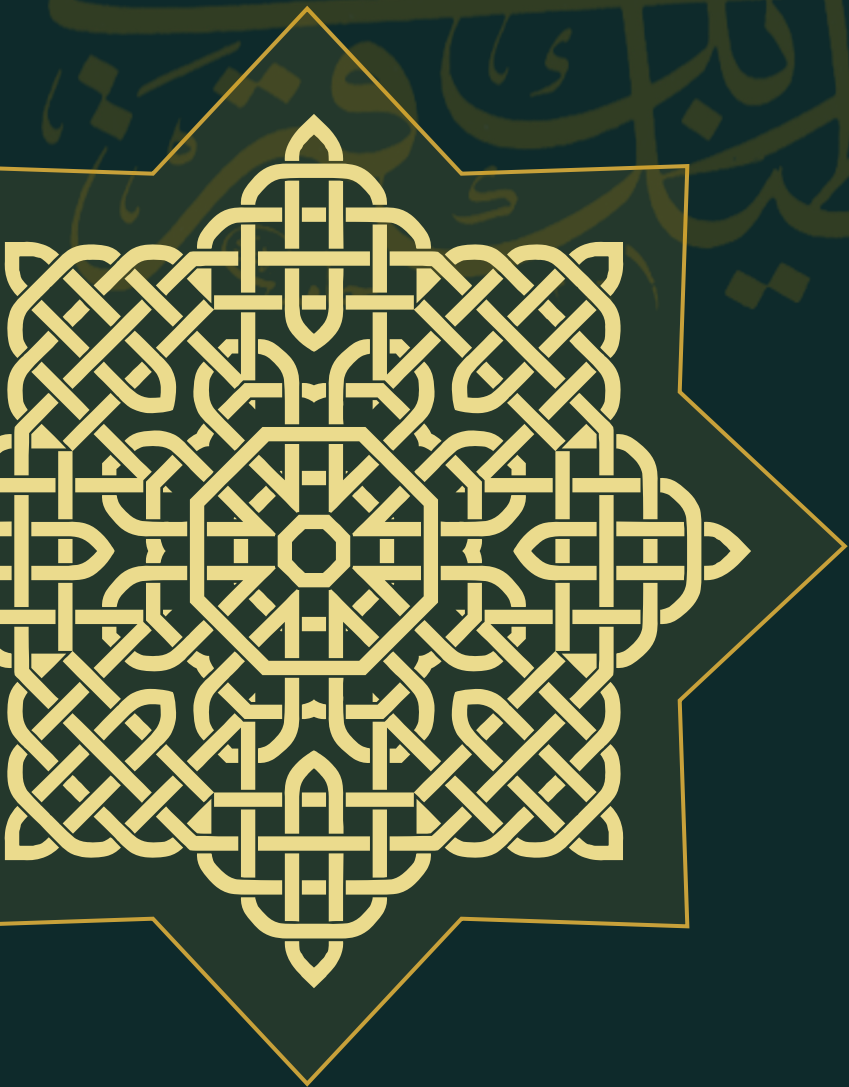


المتشابه اللفظي في القرآن الكريم

قراءة حاسوبية لبعض الروايات المتواترة
بالنماذج اللغوية الكبيرة وشبكات التعلم العميق

د. فارس بسام مسمار

أستاذ مشارك مساند في كلية الهندسة الكهربائية والحاسوب في جامعة تكساس بدالاس



نبذة عن المتحدث

مختص في مجال هندسة الحاسوب والكهرباء مع اهتمام باللغة العربية



فارس بسام مسمار قيادي تقني ومستشار. حصل على درجة البكالوريوس في هندسة الحاسوب بتقدير امتياز من الجامعة الأردنية عام ٢٠٠٤، ودرجتي الماجستير في الهندسة الكهربائية والماجستير في إدارة الأعمال بأعلى درجات التميز من جامعة تكساس في دالاس عامي ٢٠١١ و ٢٠١٤ على التوالي، ودرجة الدكتوراه في الهندسة الكهربائية وهندسة الحاسوب من جامعة تكساس في أوستن عام ٢٠١٩. وقد نال منحة مؤسسة ماركوس والنبيرغ للبحث العلمي والتعليم لعامي ٢٠١٦ و ٢٠١٧.

على مدى العقدين الماضيين، شغل الدكتور مسمار مناصب عليا وأدوارًا قيادية في شركات موتورولا وإريكسون وسامسونج ونوكيا. وقد رُفِعَ إلى عضو أول في معهد مهندسي الكهرباء والإلكترونيات عام ٢٠١٧، ونال في عام ٢٠٢٣ لقب عضو الهيئة التقنية المتميز من مختبرات نوكيا بل لايز.

يشغل الدكتور مسمار حاليًا منصب كبير المستشارين الرئيسيين في مختبرات نوكيا بل لايز منذ عام ٢٠٢٢ ويعمل أستاذًا مشاركًا مساندًا في الهندسة الكهربائية وهندسة الحاسوب بجامعة تكساس في دالاس منذ خريف عام ٢٠٢١. كما يعمل الدكتور مسمار محررًا مشاركًا لواحدة من أهم مجلات معهد مهندسي الكهرباء والإلكترونيات منذ عام ٢٠٢٥ ومحاضرًا مميزًا لجمعية الاتصالات في المعهد ذاته (٢٠٢٦-٢٠٢٧).

تشمل اهتمامات الدكتور مسمار البحثية تعلّم الآلة والذكاء الاصطناعي والاتصالات اللاسلكية الأرضية وغير الأرضية. أما في مجال اهتمامه باللغة العربية ففي الخط العربي، وكتابة الشعر، ومراجعة وتأليف مناهج اللغة العربية لغير الناطقين بها، والمحافظة على برمجيات حواسيب ٨-بت بالعربية، وإثراء مواقع الإنترنت الكبرى بالمحتوى العربي.



خطة المحاضرة

- ١ المتشابه: تعريفه وأنواعه، والبيانات المتاحة، والتحديات
- ٢ معمارية الحل المقترح ومخططه الكتلي
- ٣ مسارات الحل مع شرح لتفاصيلها
- ٤ عرض لبعض النتائج
- ٥ الخلاصة

خطة المحاضرة

المتشابه: تعريفه وأنواعه، والبيانات المتاحة، والتحديات



معمارية الحل المقترح ومخططه الكتلي



مسارات الحل مع شرح لتفاصيلها



عرض لبعض النتائج



الخلاصة



ما المقصود بالمتشابه؟

مصطلح واحد بمعنيين مختلفين، ولكل منهما أدواته

المتشابه اللفظي



آيات تتكرر ألفاظها أو تتقارب، مع فرق يسير: زيادة أو نقص، أو تقديم أو تأخير، أو إبدال كلمة أو حرف.

يعتني به الحفاظ ضبطاً، والمفسرون توجيهاً.

قال يا قوم ليس بي **ضلالة** ولكني رسول من رب العالمين (الأعراف ٦١)

قال يا قوم ليس بي **سفاهة** ولكني رسول من رب العالمين (الأعراف ٦٧)

موضوع هذه المحاضرة، ويمكن قياسه حاسوبياً

المتشابه في مقابل المحكم



المعنى الوارد في سورة آل عمران (الآية ٧): ما يحتمل أكثر من وجه أو يُردّ علمه إلى الله.

مسألة تفسيرية وعقدية، لا يُحكم فيها بنموذج حاسوبي.

(خارج نطاق القياس الآلي)

أنواع الفروق في المتشابه اللفظي

أمثلة من مخرجات البرنامج موضوع محاضرة اليوم

إبدال كلمة

... ليس بي **ضلالة** ولكني رسول ...

... ليس بي **سفاهة** ولكني رسول ...

الأعراف ٦١ ↔ الأعراف ٦٧

تقديم وتأخير

... إلا بشرى **لكم** ولتطمئن قلوبكم **به** وما النصر إلا من عند الله **العزیز الحكيم**

... بشرى ولتطمئن **به** قلوبكم وما النصر إلا من عند الله **إن الله عزيز حكيم**

آل عمران ١٢٦ ↔ الأنفال ١٠

زيادة ونقص

... وهم بالآخرة كافرون

... وهم بالآخرة **هم** كافرون

الأعراف ٤٥ ↔ هود ١٩

زيادة كلمة

... بلغ أشده آتيناها حكما وعلما ...

... بلغ أشده **واستوى** آتيناها حكما ...

يوسف ٢٢ ↔ القصص ١٤

اختلاف صيغة

وإذ **نجيناكم** من آل فرعون يسومونكم سوء العذاب **يذبحون** أبناءكم ويستحيون ...

وإذ **أنجيناكم** من آل فرعون يسومونكم سوء العذاب **يقتلون** أبناءكم ويستحيون ...

البقرة ٤٩ ↔ الأعراف ١٤١

اختلاف حرف

إنك لا تسمع ...

فإنك لا تسمع ...

النمل ٨٠ ↔ الروم ٥٢

البيانات: ثماني روايات موثقة

أربعة قرّاء بواقع راويين لكل قراءة كما في نصوص مجمع الملك فهد لطباعة المصحف الشريف



ابن كثير

آية ٦,٢٢٠

البزي

آية ٦,٢٢٠

قنبل



أبو عمرو

آية ٦,٢١٧

الدوري

آية ٦,٢١٧

السوسي



نافع

آية ٦,٢١٤

ورش

آية ٦,٢١٤

قالون



عاصم

آية ٦,٢٣٦

حفص

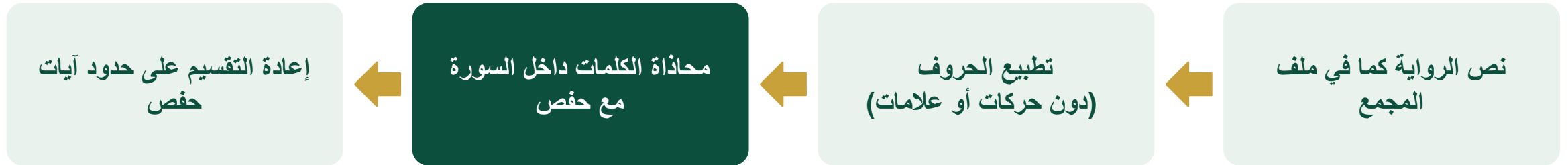
آية ٦,٢٣٦

شعبة

٨ من ٢٠ رواية متواترة (القرّاء العشرة، لكل قارئ راويان) وهي الروايات المتاحة نصًّا من المصدر المستخدم.
لاحظوا اختلاف عدد الآيات: حيث أن عدّ الآيات يختلف بين الروايات، وهذا أول تحدّي حاسوبي.

التحدي الأول: كيف نقارن روايات تختلف في عدد الآي؟

نُحاذي كل رواية على "شبكة حفص" فتصير الإحالة واحدة ويبقى نص كل رواية كما هو



مثال: أول سورة البقرة



بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ
الْحَمْدُ لِلَّهِ الَّذِي هَدَى
لِلْإِسْلَامِ أُمَّةً مِّنْهُمُ يَتَّبِعُونَ
وَالَّذِينَ هَدَى
لِلْغَيِّ ۚ إِنَّ الْغَيَّ لَا يَأْتِي
بِشَيْءٍ سَعْدٍ ۚ سَاءَ لِلَّذِينَ
أُوتُوا الْكِتَابَ مَقَرًّا يَأْتِيهِمْ
الَّذِينَ آمَنُوا يَتَّبِعُونَ



بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ
الْحَمْدُ لِلَّهِ الَّذِي هَدَى
لِلْإِسْلَامِ أُمَّةً مِّنْهُمُ يَتَّبِعُونَ
وَالَّذِينَ هَدَى
لِلْغَيِّ ۚ إِنَّ الْغَيَّ لَا يَأْتِي
بِشَيْءٍ سَعْدٍ ۚ سَاءَ لِلَّذِينَ
أُوتُوا الْكِتَابَ مَقَرًّا يَأْتِيهِمْ
الَّذِينَ آمَنُوا يَتَّبِعُونَ

المصدر: <https://easyquran.com/>

ورش: "الم ذلك الكتاب لا ريب فيه هدى للمتقين" آية واحدة (البقرة ١)

حفص: "الم" آية مستقلة، ثم "ذلك الكتاب لا ريب فيه هدى للمتقين" (البقرة ٢)

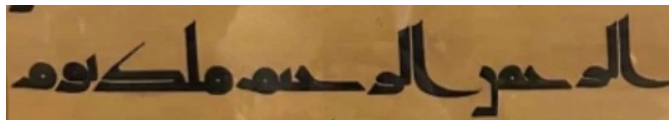
بعد المحاذاة: "الم" في موضع البقرة ١ من كل رواية، وبقية كلمات ورش تنتقل إلى موضع البقرة ٢.

شبكة حفص هي قائمة مرجعية من ٦,٢٣٦ خانة في كل خانة منها آية بترقيم رواية حفص (السورة + الآية) على العد الكوفي بهدف محاذاة الآيات عبر الروايات الأخرى.

التحدي الثاني: تعدد مستويات المقارنة بين الحروف والرسم

الرسم = هيكل الكلمة بلا نقط ولا حركات، كما كُتبت في المصاحف العثمانية. الفرش = كيفية قراءة كل كلمة مختلف فيها بين القراء.

الموضع	رواية أ	رواية ب	مستوى الحروف	مستوى الرسم
الفتحة ٤ (حفص / ورش)	مَالِكٍ	مَلِكٍ	مالك ≠ ملك	ملك = ملك *
البقرة ٧٤ (حفص / ابن كثير)	تَعْمَلُونَ	يَعْمَلُونَ	تعملون ≠ يعملون	يعملون = يعملون
الأعراف ٥٧ (حفص / الدوري)	بُشْرًا	نُشْرًا	بشرا ≠ نشرا	بسرا = بسرا
آل عمران ١٣٣ (حفص / نافع)	وَسَارِعُوا	سَارِعُوا	وسارعوا ≠ سارعوا	(يختلف الرسم)



* الكاف المتطرفة كما في المصاحف العثمانية (خط الجزم) رغم أنها مكتوبة بصورتها المعروفة اليوم (ك) في ملفات البيانات

يُسقط التطبيع أكثر الأصول عمدًا (الإمالة والتقليل، وصور الهمز، وصلة الضمير) ويُبقي فُرَش الحروف؛ لأن التشابه اللفظي يتعلق بالكلمات لا بطريقة الأداء.

خطة المحاضرة

المتشابه: تعريفه وأنواعه، والبيانات المتاحة، والتحديات



معمارية الحل المقترح ومخططه الكتلي



مسارات الحل مع شرح لتفاصيلها



عرض لبعض النتائج

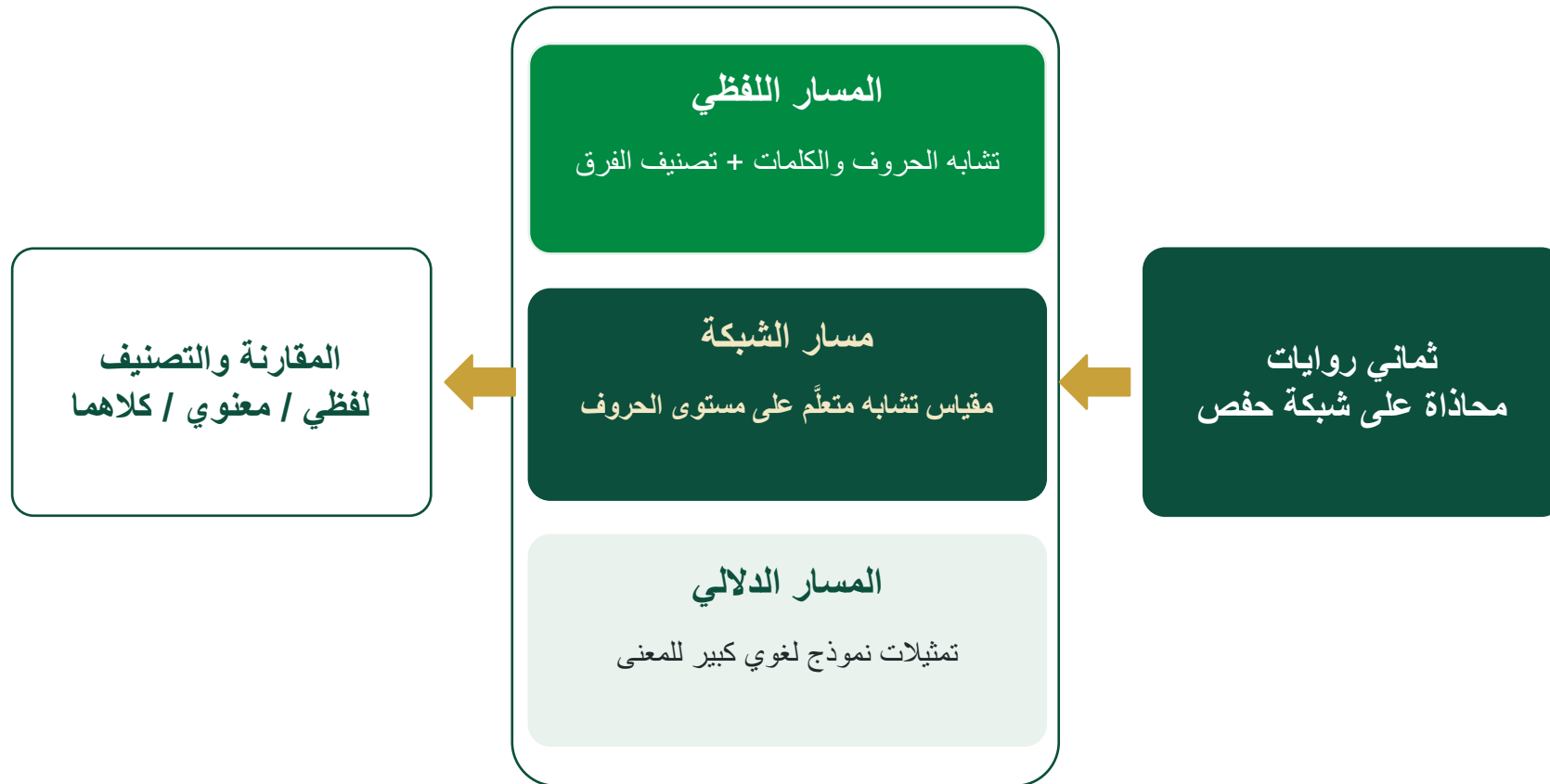


الخلاصة



معمارية البرنامج: ثلاثة مسارات ثم مقارنة

كل مسار يرى التشابه من زاوية مختلفة



ضابط أساسي: النص القرآني يُعرض دائماً من المصدر الموثق، ولا يُولد أي نموذج.



المخطط الكتلي للبرنامج: من ETL إلى المخرجات

كل كتلة تقابل دالة في البرنامج المكتوب بلغة بايثون

ETL ١

E الاستخلاص

تنزيل ٨ ملفات JSON من مجمع الملك فهد عبر GitHub



T1 التقطيع

فصل الكلمات، وحذف أرقام الآي والرموز المنفردة



T2 التطبيع

مستوى الحروف (بلا حركات، توحيد الهمز والألف) ومستوى الرسم المجرد



T3 المحاذاة

مطابقة التسلسلات مع حفص داخل كل سورة، وتوحيد حدود الكلمات



L التحميل

٨ روايات $6,236 \times$ وحدة على شبكة حفص

٢ النماذج

المسار اللفظي

TF-IDF حروف ٢-٤ + فهرس مقاطع ٥ كلمات ← تشابه ليفشتاين على الكلمات، وتداخل الكلمات، وأطول مقطع ← مصنف نوع الفرق

مسار الشبكة (BiLSTM سيامية)

تضمن حروف ٦٤ ← BiLSTM طبقتان $128 \times$ ← متوسط + أقصى ← متجه ٢٥٦ مطبّع. التدريب: InfoNCE على أزواج الروايات والمتشابه والتعزيز

المسار الدلالي (نموذج لغوي كبير)

تمثيل النص الإملاني بنموذج BAAI/bge-m3 ← عتبة المنين ٩٩,٩ من أزواج عشوائية

٣ التحليل والتقييم

فروق الروايات

مقارنة كل رواية بحفص كلمة بكلمة: اتحاد الرسم أو اختلافه

المتشابه عبر الروايات

تشغيل المسار اللفظي على كل رواية ومقارنة تصنيف الأزواج

تقييم الشبكة

R@1 و R@5 و R@10 و MRR، ومقارنة الاسترجاع بين الروايات

الأربع

لفظي ومعنوي / معنوي فقط / لفظي فقط / لا تشابه

٤ المخرجات

lafzi_pairs_<riwaya>.csv $\times 8$
exact_repeats_hafs.csv
recurring_phrases_hafs.csv
qiraat_variants.csv
cross_riwaya_mutashabihat.csv
bilstm.pt
bilstm_novel_pairs.csv
semantic_pairs.csv
llm_judgements.csv
bilstm_eval.json
summary.json

خطة المحاضرة

المتشابه: تعريفه وأنواعه، والبيانات المتاحة، والتحديات



معمارية الحل المقترح ومخططه الكتلي



مسارات الحل مع شرح لتفاصيلها



عرض لبعض النتائج



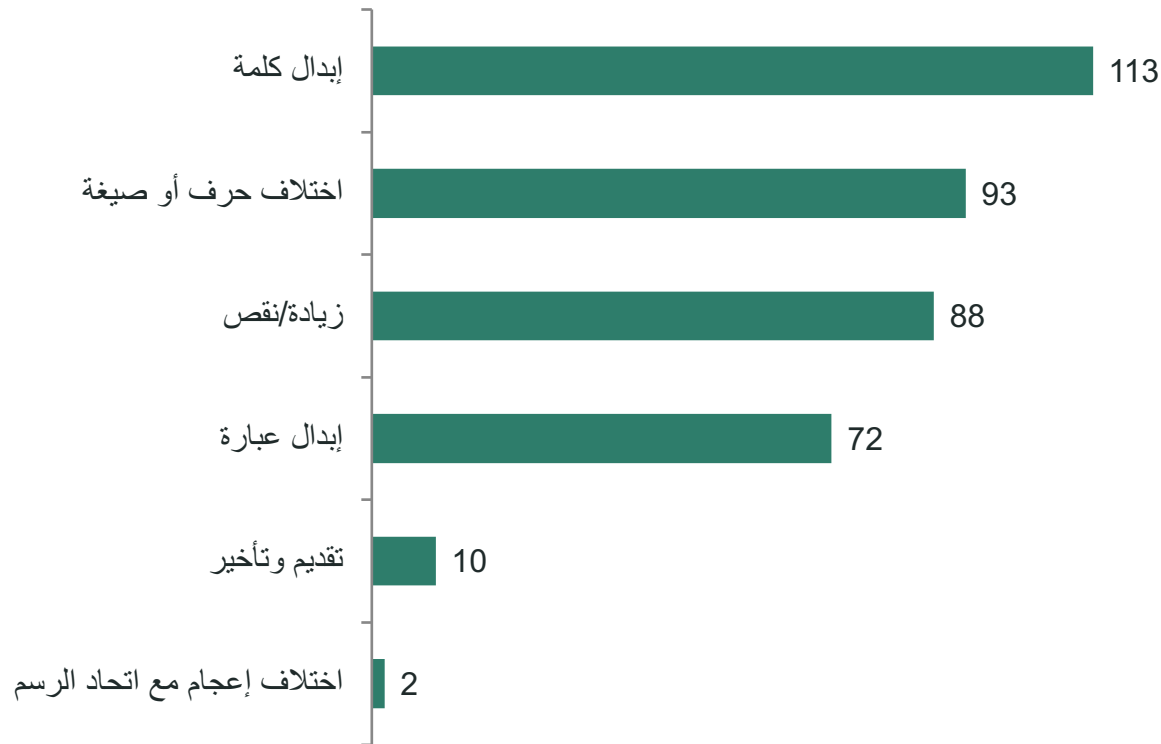
الخلاصة



المسار اللفظي: ماذا وجدنا في رواية حفص؟

المعيار: تشابه تسلسل الكلمات ٧٠٪ فأكثر، أو تداخل الكلمات ٦٥٪ فأكثر، أو مقطع مشترك من ٥ كلمات فأكثر

أنواع الفروق في الأزواج القوية



الزوج الواحد قد يجمع أكثر من نوع من هذه الفروقات

٩٨

مجموعة آيات مكررة حرفيًا
(٢٨٠ آية)

٣٠٠

زوجًا متشابهًا بقوة

٧٤٨

عبارة تتكرر في ٣ آيات فأكثر

٢,٩١٤

زوجًا يشتركان في مقطع من ٥ كلمات
فأكثر

المسار اللفظي: أمثلة من مخرجات المسار من رواية حفص

الأحمر موضع الفرق بعد المحاذاة كلمة بكلمة

إبدال كلمة

الأعراف ٦١

قال يا قوم ليس بي **ضلالة** ولكني رسول من رب العالمين

الأعراف ٦٧

قال يا قوم ليس بي **سفاهة** ولكني رسول من رب العالمين

زيادة كلمة

يوسف ٢٢

ولما بلغ أشده آتيناها حكما وعلما وكذلك نجزي المحسنين

القصص ١٤

ولما بلغ أشده **واستوى** آتيناها حكما وعلما وكذلك نجزي المحسنين

تقديم وتأخير مع زيادة

آل عمران ١٢٦

وما جعله الله إلا بشرى **لكم** ولتطمئن قلوبكم **به** وما النصر إلا من عند الله **العزیز الحكيم**

الأنفال ١٠

وما جعله الله إلا بشرى ولتطمئن قلوبكم **به** وما النصر إلا من عند الله **إن الله عزيز حكيم**

اختلاف حرف

النمل ٨٠

إنك لا تسمع الموتى ولا تسمع الصم الدعاء إذا ولوا مدبرين

الروم ٥٢

فإنك لا تسمع الموتى ولا تسمع الصم الدعاء إذا ولوا مدبرين

مسار الشبكة: BiLSTM سيامية على مستوى الحروف

تتعلم "مسافة" بين الآيات بدلاً من عتبات ثابتة



التشابه = جيب تمام الزاوية بين متجهي الآيتين

٣ تعزيز يحاكي التشابه

حذف كلمة = نقص، إدراج = زيادة، تبديل متجاورتين = تقديم وتأخير، إبدال كلمة، إبدال حرف (و/ف، ت/ي).

٢ أزواج المتشابه اللفظي

٩٧٨ زوجاً للتدريب، ويُحجب ٢٠٪ من الأزواج القوية منها (٦٠ زوجاً) للاختبار.

١ أزواج من الروايات

الموضع نفسه في روايتين مع اختلاف الألفاظ فيه (٥٦٥ موضعاً)، وهذه المواضع تصلح "تعزيزاً طبيعياً" للبيانات.

مسار الشبكة: كيف تتعلم الشبكة، وكيف نقيس أداءها؟

التعلم التبايني ذاتي الإشراف

١ الخسارة التباينية (InfoNCE) في التدريب

$$z_a = \frac{f(a)}{\|f(a)\|}, \quad z_b = \frac{f(b)}{\|f(b)\|} \quad L = \frac{1}{2} [\text{CE}(S, y) + \text{CE}(S^T, y)]$$

$$S_{ij} = \frac{z_{a,i} \cdot z_{b,j}}{\tau}, \quad y_i = i \quad \text{CE}(S, y) = -\frac{1}{B} \sum_{i=1}^B \log \frac{\exp(S_{ii})}{\sum_{j=1}^B \exp(S_{ij})}$$

a, b	آيتا الزوج الموجب: الموضع نفسه في روايتين، أو زوج متشابه، أو الآية وصورتها المعدلة
$f(\cdot)$	الشبكة: تضمين الحروف ← BiLSTM ← تجميع (متوسط + أقصى) ← طبقة خطية
z_a, z_b	متجهي الآيتين بعد التطبيع L2: طول كل منهما ١، وبُعده ٢٥٦
B	حجم الدفعة = ٩٦ زوجًا تُحسب الخسارة على الدفعة كلها معًا
S_{ij}	درجة التشابه بين الآية i في الطرف الأول والآية j في الطرف الثاني
τ	درجة الحرارة = ٠,٠٥ والقسمة عليها يُضخّم الفرق بين الدرجات
$y_i = i$	الإشارة الموجهة: نظير الآية i هو العنصر i نفسه في الطرف الآخر
CE	الإنتروبيا المتقاطعة: تصنيف من B فئة «آيتها النظير الصحيح؟»
S^T	منقول المصفوفة: الاتجاه المعاكس من b إلى a ، ومن هنا المعامل $1/2$
$j \neq i$	السلبات: بقية عناصر الدفعة، بعد إخفاء السلبات الكاذبة

٢ مقاييس الاسترجاع في التقييم

$$\text{MRR} = \frac{1}{|Q|} \sum_{q \in Q} \frac{1}{\text{rank}_q} \quad \text{R@}k = \frac{1}{|Q|} \sum_{q \in Q} \mathbb{1}[\text{rank}_q \leq k]$$

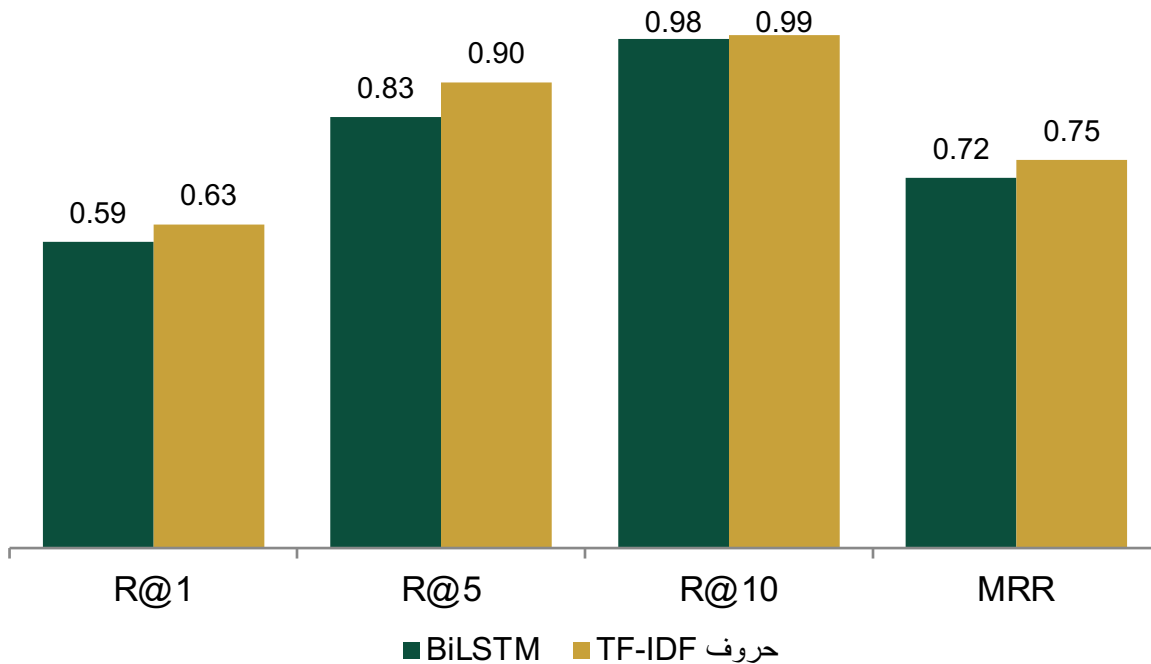
q	استعلام: آية نبحث عن نظيرها المتشابه
Q	الاستعلامات: ٦٠ زوجًا محجوبًا × اتجاهين = ١٢٠
rank_q	مرتبة النظير الصحيح بين ٦,٢٣٥ آية مرتبة بجيب التمام
$\mathbb{1}[\cdot]$	يساوي ١ إن تحقق الشرط، وصفر عدا ذلك
k	عدد النتائج الأولى المعتبرة: ١ أو ٥ أو ١٠

مثال: إن جاء النظير الصحيح ثالثاً فهو محسوب في $\text{R@}5$ و $\text{R@}10$ لا في $\text{R@}1$ ، ويُسهم بـ $1/3$ في MRR.

المقياسان بين ٠ و ١، والأعلى أفضل.

مسار الشبكة: نتائج قريبة من حيث الأداء للمسار اللفظي

الاسترجاع على ٦٠ زوجًا محجوبًا (في الاتجاهين). التدريب باستعمال ١٠ حقب على معالج رسومي



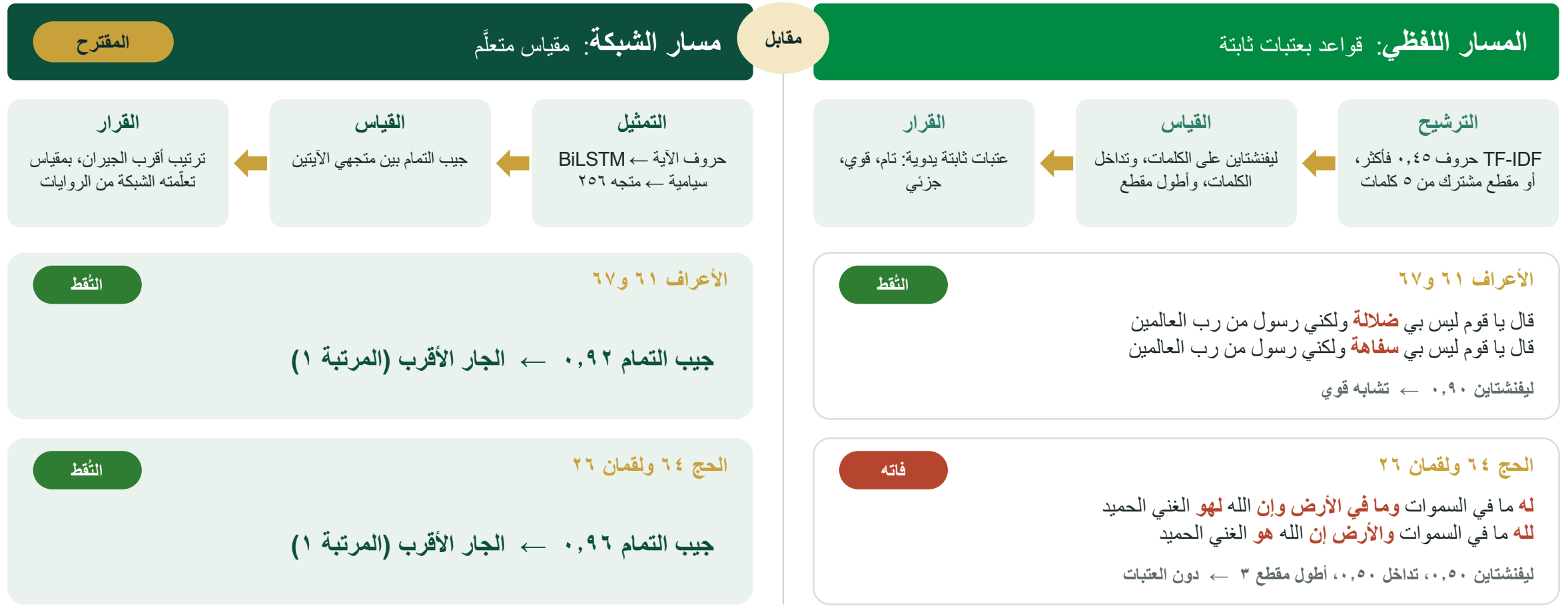
ثبات أمام اختلاف الروايات: تتعرف الشبكة على الآية بلفظ أي رواية كانت. فقد وجدت الشبكة الآية المقابلة في رواية حفص بنسبة صحة لا تقل عن ٩٨,٩٪ محسوبة لكل الروايات المتاحة (مقابل ٩٨,٥٪ لطريقة TF-IDF التقليدية).

أزواج اقترحتها الشبكة ولم يقترحها المسار اللفظي

٠,٩٦	الحج ٦٤ ↔ لقمان ٢٦
٠,٨٩	مريم ١٥ ↔ مريم ٣٣
٠,٨٩	النحل ١٢٧ ↔ النمل ٧٠
٠,٨٦	النحل ٤٠ ↔ يس ٨٢
٠,٨٦	الحجر ٤٠ ↔ الصافات ١٦٠
٠,٩١	آل عمران ١٨٩ ↔ المائدة ١٢٠
٠,٩٥	الطارق ٢ ↔ الفارعة ٣

مساران لقياس التشابه: قواعد مكتوبة ومقياس متعلم

المسار اللفظي يقرر بعتبات نضعها يدويًا ومسار الشبكة يتعلم التشابه من الروايات نفسها



التمثيل الدلالي: النموذج اللغوي الكبير كمُشَفِّر

يحوّل معنى الآية إلى متجه دون أن يكتب حرفاً من النص القرآني*



النموذج BAAI/bge-m3 (نحو ٥٦٠ مليون معامل). التشابه = جيب تمام الزاوية بين متجهي الآيتين

١ مُدَرَّب مسبقاً

نستعمل النموذج مُشَفِّراً فقط دون ضبط دقيق على ألفاظ القرآن، فيبقى المسار الدلالي مستقلاً عن تسميات المسار اللفظي.

٢ عتبة معايرة وأربع

نقارن كل آية بأقرب ٥ جيران من حيث معامل التشابه، ونعتمد عتبة المئين ٩٩,٩ من ٢٠٠ ألف زوج عشوائي، ثم نضع كل زوج في ربع: لفظي ومعنوي، معنوي فقط، لفظي فقط، لا تشابه.

* لا يولّد النموذج هنا نصّاً قرآنياً، ولا يحكم في التشابه اللفظي ولا يوجّه ولا يرجّح. دوره التمثيل (عبر التضمين) والتصنيف فقط.

التمثيل الدلالي: أمثلة من مخرجات المسار

قام التمثيل الدلالي بعد التضمين بحساب عتبة التشابه المعنوي عند ٠,٦٩ (= المئين ٩٩,٩ من ٢٠٠ ألف زوج عشوائي)

لفظي فقط

٧ زوجًا

الصفات ١٣٣ ↔ الصفات ١٣٩

وإن **لوطا** لمن المرسلين
وإن **يونس** لمن المرسلين

معنوي ٠,٦٨، حرفي ٠,٦١

المطففين ١٩ ↔ الهمزة ٥

وما أدراك ما **عليون**
وما أدراك ما **الحطمة**

معنوي ٠,٦٥، حرفي ٠,٤٩

معنوي فقط

٩,٥٤٧ زوجًا

التكوير ٢ ↔ الانفطار ٢

وإذا النجوم انكدرت
وإذا الكواكب انتثرت

معنوي ٠,٨٢، حرفي ٠,١٤

الطور ٧ ↔ البروج ١٢

إن عذاب ربك لواقع
إن بطش ربك لشديد

معنوي ٠,٨٠، حرفي ٠,١٥

النحل ١٩ ↔ الانشقاق ٢٣

والله يعلم ما تسرون وما تعلنون
والله أعلم بما يوعون

معنوي ٠,٨٦، حرفي ٠,١٨

لفظي ومعنوي

١,٧٩١ زوجًا

الأعراف ٦١ ↔ الأعراف ٦٧

قال يا قوم ليس بي **ضلالة** ولكني رسول من رب العالمين
قال يا قوم ليس بي **سفاهة** ولكني رسول من رب العالمين

معنوي ٠,٩٧، حرفي ٠,٧٥

يوسف ٢٢ ↔ القصص ١٤

ولما بلغ أشده آتيناها حكما وعلمنا وكذلك نجزي المحسنين
ولما بلغ أشده **واستوى** آتيناها حكما وعلمنا وكذلك نجزي
المحسنين

معنوي ٠,٩٦، حرفي ٠,٩٣

يمكن تدريب النموذج اللغوي الكبير تدريبًا إضافيًا لضبط وتحسين مخرجات المسار بعد معاينة جميع المخرجات. وهذا لا غنى فيه عن علم المختصين، وهو خارج نطاق المحاضرة.

التشغيل والإعدادات: الأوامر والمعاملات والبنية

المرجع التقني للبرنامج والقيم الافتراضية كما في لغة بايثون

١ التشغيل من Terminal

البند	الأمر أو الخيار
التثبيت	<code>pip3 install numpy scikit-learn rapidfuzz torch sentence-transformers</code>
تشغيل كامل	<code>python3 quran_mutashabihat.py --out out</code>
الروايات	<code>--riwayat all hafs,warsh,...</code>
حقب BiLSTM	<code>--epochs 10</code>
نموذج التمثيل	<code>--semantic-model BAAI/bge-m3 lsa</code>
تخطي مسار	<code>--no-bilstm --no-semantic</code>
التخزين والبذرة	<code>--cache data_kfgqpc --seed 7</code>
العتاد	يُكتشف المعالج الرسمي تلقائيًا، وإلا فالمعالج المركزي

٢ معاملات تدريب BiLSTM

البند	القيمة
المُدخل	حروف. المفردات ٣٢ (٣٠ حرفاً + حشو + مجهول)
التضمين	٦٤ بُعدًا
BiLSTM	طبقتان ١٢٨ × لكل اتجاه، إسقاط ٠,١
التجميع	متوسط + أقصى (٥١٢) ← خطي ٢٥٦
المعاملات	٧٢٧,٢٩٦
الخسارة	InfoNCE متماثلة، $\tau = ٠,٠٥$ ، إخفاء السلبيات الكاذبة
المُحسِّن	AdamW معدل ٠,٠٠٢، اضمحلال ٠,٠٠٠١
التدريب	دفعة ٩٦ × ٦٤ خطوة × ١٠ حقب = ٦٤٠ خطوة
الطول الأقصى	٢٠٠ حرف تدريبًا، ٦٠٠ استدلالًا
مزيج العينات	٣٥٪ متشابه لفظي، ٢٥٪ روايات، ٤٠٪ تعزيز
الاختبار	٢٠٪ من الأزواج القوية (٦٠ زوجًا)، البذرة ٧

٣ بنية النموذج اللغوي

البند	القيمة
الدور	تمثيل دلالي (Embedding) فقط، لا توليد
النموذج	BAAI/bge-m3
الأساس	XLM-RoBERTa-large كمُشَفِّر
الطبقات	٢٤ طبقة × ١٦ رأس انتباه
البُعد المخفي	١,٠٢٤ (وهو بُعد متجه الآية)
المعاملات	نحو ٥٦٠ مليونًا
المفردات	٢٥٠,٠٠٢ وحدة؛ أكثر من ١٠٠ لغة
أقصى طول	٨,١٩٢ وحدة
التجميع	متجه [CLS] مطبَّع (تمثيل كثيف)
الإدخال	نص حفص الإملاني، آيات من ٣ كلمات فأكثر
العتبة	المنين ٩٩,٩ من ٢٠٠ ألف زوج عشوائي مع أقرب ٥ جيران

حدود المنهج وضوابطه

ما يستطيعه الحاسوب، وما يبقى لأهل العلم والاختصاص

التغطية

ثمانين روايات من عشرين متواترة والبرنامج مصمم لإضافة البقية متى توفر نصها الموثق.

العتبات

حدود "القوي" و"الجزئي" اختيارية، فالأعداد تتغير بتغيرها، والمنهج ثابت وقابل للتكرار.

التطبيع

يُسقط أكثر الأصول (الإمالة وصور الهمز والصلات) ويُبقي فرش الحروف، وهو اختيار منهجي.

التسميات

أزواج التدريب مستمدة جزئياً من المسار اللفظي. الفیصل (المعيار الذهبي) من كتب المتشابه المعروفة.

الإحالة

بترقيم حفص (العد الكوفي) لتوحيد المقارنة، مع عرض نص كل رواية كما هو.

الدور

البرنامج يحصي ويصنّف ويقترح وأما التوجيه والترجيح فمتروك لأهل الاختصاص.

خطة المحاضرة

المتشابه: تعريفه وأنواعه، والبيانات المتاحة، والتحديات



معمارية الحل المقترح ومخططه الكتلي



مسارات الحل مع شرح لتفاصيلها



عرض لبعض النتائج



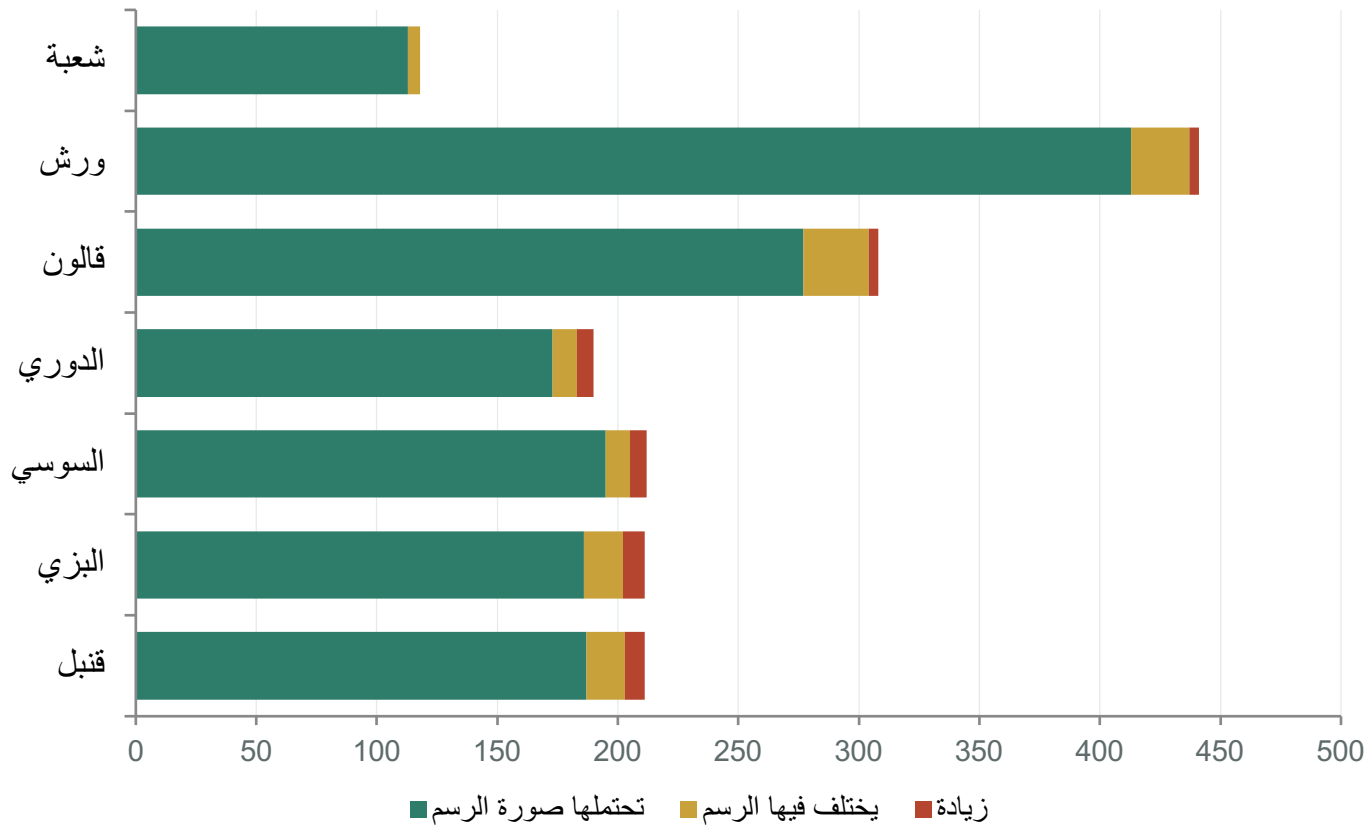
الخلاصة



النتائج: كيف تختلف الروايات المتواترة عن رواية حفص؟

فروق الكلمات بعد إسقاط أكثر الأصول

عدد فروق الكلمات مقارنة برواية حفص



٩٣%

من فروق الكلمة الواحدة تحتملها صورة الرسم نفسها
(تختلف في النقط والشكل فقط)

وما يختلف فيه الرسم يوافق رسم المصاحف العثمانية المختلفة،
ومنه:

وسار عوا / سار عوا: آل عمران ١٣٣ (قراءة نافع بروايتيه)
ووصى / وأوصى: البقرة ١٣٢ (قراءة نافع بروايتيه)
تحتها / من تحتها: التوبة ١٠٠ (قراءة ابن كثير بروايتيه)
فيما / بما: الشورى ٣٠ (قراءة نافع بروايتيه)

النتائج: "بغافل عما تعملون / يعملون"

كل مواضع "وما لله / وما ربك بغافل عما..." : الفاصلة تتبدل بين الخطاب والغيبة بحسب الرواية

الآية	حفص	شعبة	ورش	قالون	الدوري	السوسي	اليزي	قتل
البقرة ٧٤	تعملون	تعملون	تعملون	تعملون	تعملون	تعملون	يعملون	يعملون
البقرة ٨٥	تعملون	يعملون	يعملون	يعملون	تعملون	تعملون	يعملون	يعملون
البقرة ١٤٠	تعملون	تعملون	تعملون	تعملون	تعملون	تعملون	تعملون	تعملون
البقرة ١٤٤	يعملون	يعملون	يعملون	يعملون	يعملون	يعملون	يعملون	يعملون
البقرة ١٤٩	تعملون	تعملون	تعملون	تعملون	يعملون	يعملون	تعملون	تعملون
آل عمران ٩٩	تعملون	تعملون	تعملون	تعملون	تعملون	تعملون	تعملون	تعملون
الأنعام ١٣٢	يعملون	يعملون	يعملون	يعملون	يعملون	يعملون	يعملون	يعملون
هود ١٢٣	تعملون	يعملون	تعملون	تعملون	يعملون	يعملون	يعملون	يعملون
النمل ٩٣	تعملون	يعملون	تعملون	تعملون	يعملون	يعملون	يعملون	يعملون

في حفص: البقرة ٧٤ و ٨٥ "تعملون" فهما متشابهتان في الفاصلة. عند نافع (بروايته) وشعبة: ٨٥ "يعملون" (مختلفتان). عند ابن كثير: كلتاها "يعملون" (متشابهتان).

النتائج: المتشابه نفسه يتغير الرواية

أزواج يختلف حكم تشابهها من رواية إلى أخرى

١ المؤمنون ٨٥ و ٨٧ و ٨٩

"سيقولون لله" ثلاث مرات في حفص، وعند أبي عمرو: "لله" ثم "الله" "الله". فيضعف التشابه بين ٨٥ و ٨٧ عند أبي عمرو (بروايته).

٢ الحديد ٢٤ والممتحنة ٦

"فإن الله هو الغني الحميد" مقطع مشترك في حفص، وعند نافع (بروايته) "فإن الله الغني الحميد" بلا "هو"، ولا يعود المقطع مشتركاً.

٣ البقرة ٧٤ و ٨٥

الفاصلة مشتركة بين البقرة ٧٤ و ٨٥ في حفص وأبي عمرو (يعملون) وعند ابن كثير بروايته (يعملون). أما عند شعبة ونافع فالفواصل مختلفة (يعملون / يعملون).

٤ التوبة ١٠٠

"جنات تجري تحتها الأنهار" في حفص، وعند ابن كثير "تجري من تحتها الأنهار"، فتتضم إلى عائلة هذه الفاصلة ويظهر ٢٣ زوجاً متشابهاً جديداً.

٥١

زوجاً يتغير تصنيفه (قوي / جزئي /
غير متشابه) بحسب الرواية

خطة المحاضرة

المتشابه: تعريفه وأنواعه، والبيانات المتاحة، والتحديات



معمارية الحل المقترح ومخططه الكتلي



مسارات الحل مع شرح لتفاصيلها



عرض لبعض النتائج



الخلاصة



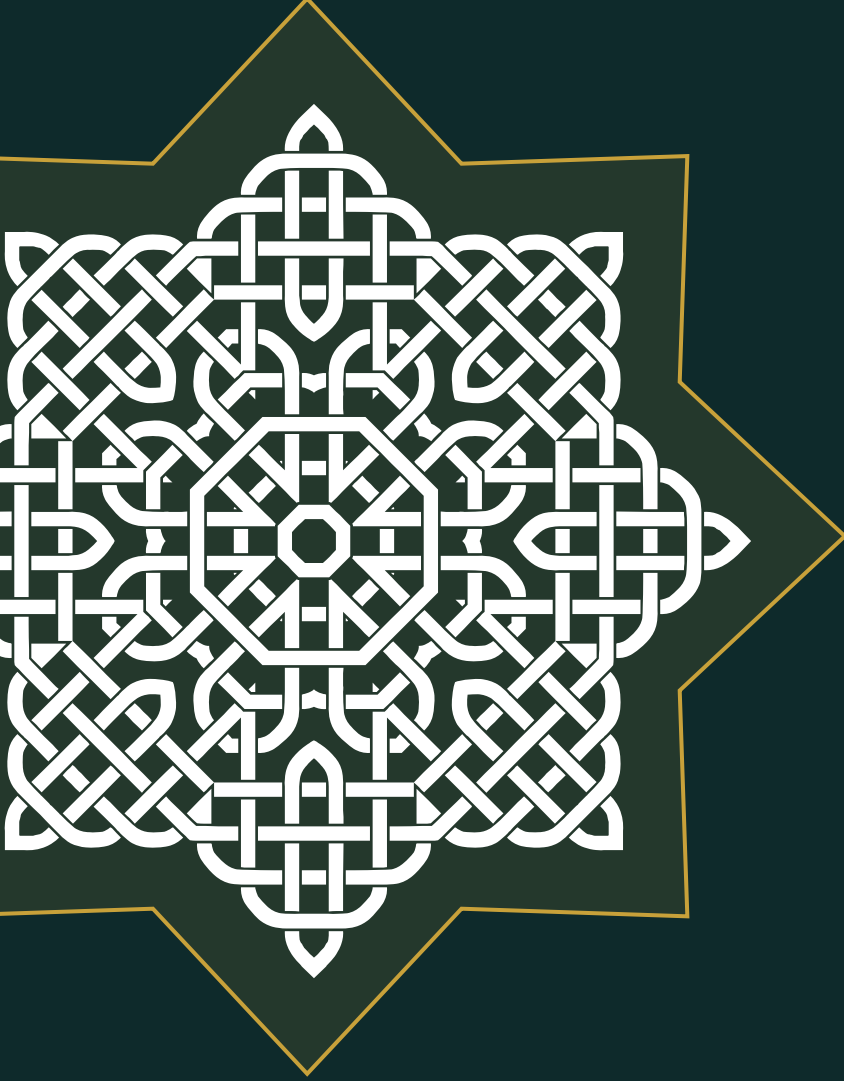
خلاصة

١ المتشابه اللفظي يمكن حصره وتصنيفه حاسوبياً وفق معايير تشغيلية قابلة للتكرار.

٢ المحاذاة على شبكة حفص تجعل الروايات المتواترة قابلة للمقارنة كلمة بكلمة.

٣ خريطة المتشابهات تتغير بتغير الرواية، وهذا بُعد جديد للحفاظ والباحثين.

٤ القراءات المتواترة مصدر بيانات تدريب موثّق للشبكات العصبية.



شكرًا لحسن استماعكم!

ملحق: مسرد المصطلحات (١ من ٢)

المصطلحات التقنية الواردة في المحاضرة والبرنامج ومقابلها الإنجليزي

المسار اللفظي

Lexical pipeline	المسار اللفظي
Character n-gram TF-IDF	TF-IDF حروف
Word-level normalized Levenshtein similarity	تشابه تسلسل الكلمات
Bag-of-words overlap	تداخل الكلمات
Shared n-gram (shingle)	مقطع مشترك
Inverted index	فهرس المقاطع
Diff-type classifier	مصنّف نوع الفرق
Threshold	العتبة

التقييم

Retrieval	الاسترجاع
Held-out test set	مجموعة الاختبار المحجوبة
Robustness	المتانة
Gold standard (ground truth)	المعيار الذهبي
Quadrants	الأرباع

البيانات ومعالجتها (ETL)

Extract / Transform / Load (ETL)	الاستخلاص / التحويل / التحميل
Tokenization	التقطيع
Text normalization	التطبيع
Sequence alignment	المحاذاة
Hafs reference grid (canonical index)	شبكة حفص
Canonical referencing / indexing	الإحالة
Character level	مستوى الحروف
Consonantal skeleton (rasm)	الرسم المجرد
Diacritical dotting (i'jām)	الإعجام
Same / different consonantal skeleton	اتحاد الرسم / اختلاف الرسم
Reproducibility	قابلية التكرار
Random seed	البذرة

ملحق: مسرد المصطلحات (٢ من ٢)

المصطلحات التقنية الواردة في المحاضرة والبرنامج ومقابلها الإنجليزي

النموذج اللغوي والمسار الدلالي

Large language model (LLM)	النموذج اللغوي الكبير
Semantic pipeline	المسار الدلالي
Semantic embedding / embedding vector	التمثيل الدلالي / متجه التمثيل
Encoder	المُشفّر
Transformer	المحوّل
Attention head	رأس الانتباه
Hidden dimension	البُعد المخفي
Vocabulary / token	المفردات / الوحدة
Maximum sequence (context) length	أقصى طول
Dense representation	التمثيل الكثيف
Nearest neighbors	أقرب الجيران
Percentile	المنين
Calibrated threshold	العتبة المعاييرة
Closed label set	قائمة مغلقة

تمت الترجمة من اللغة الإنجليزية إلى العربية باستعمال مساعد الذكاء الاصطناعي Claude.

الشبكة العصبية (BiLSTM)

Siamese network	شبكة سيامية
Bidirectional LSTM (BiLSTM)	ذاكرة طويلة قصيرة المدى ثنائية الاتجاه
Learned similarity metric (metric learning)	مقياس تشابه متعلّم
Embedding (layer)	التضمين
Pooling (mean + max)	التجميع (متوسط + أقصى)
L2-normalized vector	متجه مطبّع
Cosine similarity	تشابه جيب تمام الزاوية
Contrastive loss (InfoNCE)	الخسارة التباينية
Temperature	درجة الحرارة T
False negatives / masking	السلبات الكاذبة / إخفاؤها
Data augmentation	تعزيز البيانات
Epoch / batch / step	الحقبة / الدفعة / الخطوة
Optimizer	المُحسّن
Learning rate	معدل التعلم
Weight decay	اضمحلال الأوزان
Gradient clipping	قص التدرج
Dropout	الإسقاط
Parameters	المعاملات
Input / inference	المُدخل / الاستدلال

ملحق: جدول للنتائج التفصيلية (أفضل ن = ٦ نتائج)

لفظي

الآية أ	الآية ب	نوع الفرق	تسلسل	تداخل	مقطع	حرفي	نص الآية أ	نص الآية ب
الحج ٦٢	لقمان ٣٠	زيادة/نقص	٠,٩٤	٠,٩٤	١٠	١,٠٠	ذلك بأن الله هو الحق وأن ...	ذلك بأن الله هو الحق وأن ...
النحل ٤٣	الأنبياء ٧	زيادة/نقص	٠,٩٣	٠,٩٣	١٢	١,٠٠	وما أرسلنا من قبلك إلا رجالا ...	وما أرسلنا قبلك إلا رجالا نوحى ...
الروم ٣٧	الزمر ٥٢	إبدال كلمة	٠,٩٣	٠,٩٣	١٣	٠,٩٣	أو لم يروا أن الله يبسط ...	أو لم يعلموا أن الله يبسط ...
النمل ٨١	الروم ٥٣	اختلاف حرف أو صيغة	٠,٩٣	٠,٩٣	١١	٠,٩٦	وما أنت بهادي العمى عن ضلالتهم ...	وما أنت بهاد العمى عن ضلالتهم ...
الأعراف ٤٥	هود ١٩	زيادة/نقص	٠,٩١	٠,٩١	٩	٠,٩٩	الذين يصدون عن سبيل الله ويبغونها ...	الذين يصدون عن سبيل الله ويبغونها ...
النمل ٨٠	الروم ٥٢	اختلاف حرف أو صيغة	٠,٩١	٠,٩١	١٠	٠,٩٨	إنك لا تسمع الموتى ولا تسمع ...	فإنك لا تسمع الموتى ولا تسمع ...

lafzi_pairs_hafs.csv

شبكي

الآية أ	الآية ب	جيب التمام	نص الآية أ	نص الآية ب
الحج ٦٤	لقمان ٢٦	٠,٩٦	له ما في السموات وما في الأرض وإن الله ...	لله ما في السموات والأرض إن الله هو الغنى ...
آل عمران ١٨٩	المائدة ١٢٠	٠,٩١	ولله ملك السموات والأرض والله على كل شيء قدير	لله ملك السموات والأرض وما فيهن وهو على كل ...
مريم ١٥	مريم ٣٣	٠,٨٩	وسلام عليه يوم ولد ويوم يموت ويوم يبعث حيا	والسلام علي يوم ولدت ويوم أموت ويوم أبعث حيا
النحل ٤٠	يس ٨٢	٠,٨٦	إنما قولنا لشيء إذا أردناه أن نقول له كن ...	إنما أمره إذا أراد شيئا أن يقول له كن ...
المرسلات ١٤	المطففين ١٩	٠,٨٦	وما أدراك ما يوم الفصل	وما أدراك ما عليون
الأنبياء ٧٠	الصفات ٩٨	٠,٨٥	وأرادوا به كيدا فجعلناهم الأخسرين	فأرادوا به كيدا فجعلناهم الأسفلين

bilstm_novel_pairs.csv

دلالي

الآية أ	الآية ب	معنوي	حرفي	الربع	نص الآية أ	نص الآية ب
الحج ٦٢	لقمان ٣٠	١,٠٠	١,٠٠	لفظي ومعنوي	ذلك بأن الله هو الحق وأن ما ...	ذلك بأن الله هو الحق وأن ما ...
المؤمنون ٨٣	النمل ٦٨	١,٠٠	١,٠٠	لفظي ومعنوي	لقد وعدنا نحن وآباؤنا هذا من قبل ...	لقد وعدنا نحن وآباؤنا من قبل ...
النحل ٤٣	الأنبياء ٧	٠,٩٩	١,٠٠	لفظي ومعنوي	وما أرسلنا من قبلك إلا رجالا نوحى ...	وما أرسلنا قبلك إلا رجالا نوحى إليهم ...
الأنعام ١١٧	القلم ٧	٠,٩٩	٠,٨٩	لفظي ومعنوي	إن ربك هو أعلم من يضل عن ...	إن ربك هو أعلم بمن ضل عن ...
الانشقاق ٢٥	التين ٦	٠,٩٩	٠,٩٤	لفظي ومعنوي	إلا الذين آمنوا وعملوا الصالحات لهم أجر ...	إلا الذين آمنوا وعملوا الصالحات فلهم أجر ...
النمل ٨٠	الروم ٥٢	٠,٩٩	٠,٩٨	لفظي ومعنوي	إنك لا تسمع الموتى ولا تسمع الصم ...	فإنك لا تسمع الموتى ولا تسمع الصم ...

semantic_pairs.csv