

آفاق اللسانيات الحاسوبية في استعادة التراث العربي: من التعرف الضوئي إلى التحليل الدلالي.

Horizons of Computational Linguistics in Restoring Arabic Heritage: From Optical Character Recognition to Semantic Analysis

فتيحة غراوي

عبد الغني بن أحمد

المركز الجامعي بركة طريق امڤوكال بركة 05001 الجزائر
مخبر: دراسات في اللغة والأدب . مخبر: دراسات في اللغة والأدب

gheraoui.fatiha@cu-barika.dz

abdelghani.benahmed@cu-barika.dz

تاريخ النشر: 2026-06-09

تاريخ القبول: 2026-05-01

تاريخ الاستلام: 2026-03-30

ملخص:

تتناول هذه الورقة البحثية إشكالية رقمنة تراث لغوي عربي في ظل فجوة عميقة بين وفرة مادة خام وقصور أدوات معالجة آلية ذكية، مما يسبب اغتراباً رقمياً لنصوص تراثية حبيسة صيغ تصويرية صامتة. تهدف دراسة إلى تأصيل انتقال مفاهيمي وتقني من حفظ وثائقي تقليدي إلى استثمار معرفي يحول نصوصاً إلى بيانات مهيكلة قابلة للفهم والتحليل، وتكمن أهمية بحث في صون هوية عربية ودمجها ضمن اقتصاد معرفة عالمي. اعتمدت دراسة منهجاً وصفيّاً تحليلياً لرصد واقع تقني، مستخدمة أدوات لسانية حاسوبية وخوارزميات معالجة لغات طبيعية وذخائر لغوية معيارية. خلصت نتائج إلى أن رقمنة ذكية تمثل مشروع أمن قومي معرفي يتجاوز أرشفة تقليدية، مؤكدة ضرورة بناء فرق بحثية بينية تجمع لغويين وتقنيين لردم هوة معرفية قائمة. كما أوصت دراسة بتأسيس هيئة وطنية موحدة لبناء مدونات نصية ضخمة ومحقة لتكون وقوداً لتدريب نماذج ذكاء اصطناعي، بما يضمن خلوداً رقمياً لتراث لغوي عابر لزمان ومكان. **الكلمات المفتاحية:** تراث لغوي، رقمنة، معالجة لغات طبيعية، لسانيات حاسوبية، ذكاء اصطناعي، ذخيرة لغوية.

Abstract :

This research paper addresses the problem of digitizing Arabic linguistic heritage within the context of a profound gap between abundant raw material and the insufficiency of intelligent automated processing tools. This gap leads to the digital alienation of heritage texts trapped in silent image-based formats. The study aims to conceptualize a conceptual and technical shift from traditional documentary preservation to cognitive investment that transforms texts into structured, analyzable data. The importance of this research lies in safeguarding Arabic identity and integrating it into the global knowledge economy. Adopting a descriptive-analytical methodology, the study examines the current technical reality using computational linguistics tools, natural language processing (NLP) algorithms, and standard linguistic corpora. The results conclude that smart digitization is a national knowledge-security project that transcends traditional archiving. It emphasizes the necessity of interdisciplinary research teams—comprising linguists and

technologists—to bridge the existing knowledge gap. The study recommends establishing a national entity to build massive, verified text corpora to serve as fuel for training artificial intelligence models, ensuring digital immortality for linguistic heritage across time and space.

Keywords:

Linguistic heritage, digitization, natural language processing, computational linguistics, artificial intelligence, linguistic corpus

*عبد الغني بن أحمد

مقدمة:

يُعد التراث اللغوي العربي واحداً من أضخم المستودعات المعرفية وأكثرها ثراءً في خريطة الحضارة الإنسانية قاطبة؛ فهو ليس مجرد ركام من النصوص التاريخية، بل هو كيان حي يحتزل في مدوناته ومخطوطاته "الذاكرة العصبية" لأمة صاغت على مدار قرون طويلة أدق علوم الآلة (من نحو وصرف وبلاغة) وأعقد علوم الغاية (من فلسفة وفقه وطب وفلك). إن هذا الإرث، الذي يمتد جغرافياً من أقاصي الشرق إلى الأندلس، يمثل نظاماً معرفياً متكاملًا (Dynamic Knowledge System) أسهم في تشكيل العقل البشري، وظل صامداً أمام عوادي الزمن بفضل عبقرية لغته ومرونتها الاشتقاقية الفريدة.

ومع تسارع وتيرة الثورة الصناعية الرابعة وانفجار عصر المعلومات، لم يعد النظر إلى هذا التراث بوصفه مجرد وثائق أرشيفية ميتة معدة للحفظ المتحفّي خياراً مقبولاً، بل استحال في المنظور المعاصر إلى "بيانات ضخمة (Big Data)" تمتلك قيمة استراتيجية كبرى. إن هذا التحول يقتضي استنطاقاً تقنياً عميقاً يهدف إلى دمج هذا المخزون اللساني في نسيج "اقتصاد المعرفة" العالمي، حيث أصبحت اللغة هي المادة الخام والمورد الأساسي للصناعات الذكية. بيد أن هذا الزخم المعرفي يواجه اليوم تحدياً وجودياً حاسماً يتمثل في الفجوة السحيقة بين "وفرة المادة الخام" المتناثرة في المكتبات العالمية، و"قصور أدوات المعالجة الآلية الذكية" ¹¹ القادرة على فك شفرات هذه النصوص وتوظيفها تقنياً.

وتتبلور معضلة البحث الحقيقية في أن أغلب الجهود الراهنة لا تزال ارتحاناً لما يمكن تسميته بـ "الرقمنة الساكنة" أو "الرقمنة الصامتة"، التي تكتفي بتحويل الورق المخطوط إلى صور رقمية جامدة (Scanned Images) ورغم أهمية هذه الخطوة في الحماية المادية، إلا أنها تظل عاجزة عن تحرير النص من قيد "الصورة" إلى فضاء "البيانات" المهيكل القابلة للقراءة الآلية، مما يحرم الحاسوب من القدرة على النفاذ إلى المحتوى، ويجعل البحث داخل ملايين الصفحات التراثية أمراً مستحيلاً بالمعايير التقنية الحديثة.

ومن هذا المنطلق، تنبثق إشكالية هذا البحث لتطرح تساؤلاً جوهرياً ومحورياً: كيف يمكننا الانتقال بالتراث العربي من مرحلة "الحفظ الوثائقي" التقليدي القائم على الأرشفة البصرية، إلى مرحلة "الاستثمار المعرفي" القائم على الفهم الآلي؟ وما هي الآفاق التقنية والمنهجية التي تفتحها اللسانيات الحاسوبية بوصفها الأداة المركزية لإعادة هندسة هذا التراث وتحويله من إرث جامد إلى مورد رقمي فاعل؟

وتأسيساً على هذه الإشكالية، يسعى البحث إلى اختبار فرضيتين أساسيتين تدعمان مساره:

الفرضية الأولى: تذهب إلى أن تعثر مشاريع الرقمنة العربية المعاصرة في استيعاب التراث يعود بالأساس إلى خلل منهجي يتعامل مع النص التراثي ككيان "بصري (Object)"، متجاهلاً طبيعته كبيانات "لسانية" معقدة تحتاج إلى توصيف صرفي ونحوي ودلالي دقيق قبل برمجتها.

الفرضية الثانية: ترى أن تبني خوارزميات المعالجة الآلية للغات الطبيعية (NLP) المتقدمة، وتوظيف تقنيات التعلم العميق، يمثل "الجسر السيرياني" الوحيد القادر على نقل التراث العربي من عزلته التقنية الحالية إلى الفاعلية الحضارية، مما يتيح للأجيال الجديدة (المواطنين الرقميين) استعادة تراثهم بلغة العصر.

ولتحقيق هذه الغايات العلمية، اعتمدت الدراسة المنهج الوصفي التحليلي؛ كونه الأداة الأنسب لرصد واقع الرقمنة العربي الراهن وتفكيك مشكلاته، مع تحليل التقنيات الحاسوبية المعاصرة ومواءمتها مع خصوصية الحرف العربي والنظام الاشتقاقي للغة الضاد. وقد استعانت الدراسة في ذلك بأدوات تقنية ولسانية متداخلة التخصصات، تشمل مفاهيم اللسانيات الحاسوبية (Computational Linguistics)، وتطبيقات الذكاء الاصطناعي في التعرف على الكتابة اليدوية (HTR)، ومعايير بناء الدخائر اللغوية المعيارية (Gold Standard Corpora)، وصولاً إلى صياغة رؤية استشرافية توازن بانسجام بين صرامة القاعدة النحوية التراثية وذكاء الخوارزمية البرمجية الحديثة.

I - جدلية المفهوم: بين الرقمنة "الساكنة" والمعالجة "الذكية":

إن الخوض في غمار هذا الموضوع يقتضي أولاً تحرير المصطلحات لضبط مسار البحث. فكثيراً ما يحدث خلطٌ منهجي بين مفهوم "الرقمنة (Digitization)" ومفهوم "المعالجة (Processing)" يشير المفهوم الأول في أدبيات علم المعلومات إلى العملية التقنية التي يتم بموجبها تحويل مصادر المعلومات التقليدية الورقية إلى صيغ إلكترونية ثنائية (Binary) قابلة للتخزين في الوسائط الحاسوبية¹ (عبد الرحمن، 2010، ص. 45). وفي سياق

¹ - عبد الرحمن، فرج. (2010) الرقمنة في المكتبات ومراكز المعلومات العربية. القاهرة: دار الفكر العربي، ص 45.

التراث، غالباً ما تقف هذه العملية عند حدود "الرقمنة التصويرية"، أي تحويل المخطوطة إلى صورة رقمية (Image)، وهي خطوة - وإن كانت ضرورية للحفظ المادي - إلا أنها تُبقي النص في حالة "صمت رقمي"، حيث لا تستطيع الآلة قراءة محتواه أو البحث في داخله.

أما المعالجة الآلية، فهي المرحلة المتقدمة التي تلي الرقمنة، وتندرج ضمن حقل "اللسانيات الحاسوبية". وتعرف بأنها إخضاع النص الرقمي لعمليات خوارزمية تهدف إلى الفهم، والتحليل، والتركيب، واستنباط المعلومات² (علي، 1988، ص. 112).

فإذا كانت الرقمنة تمنحنا "صورة" الكلمة، فإن المعالجة تمنحنا "هوية" الكلمة (جذرها، وزنها، وظيفتها النحوية). وعليه، فإن الرهان الحقيقي اليوم ليس في تكديس صور المخطوطات على الخوادم (Servers)، بل في تحويل تلك الصور إلى نصوص رمزية (Text) يمكن للحاسوب "فهمها" ومعالجتها، وهو ما ينقلنا من مرحلة "الحفظ المتحفي" إلى مرحلة "الاستثمار المعرفي".

II - الأهمية الاستراتيجية لرقمنة التراث اللغوي:

تكتسب عملية رقمنة التراث اللغوي العربي ومعالجته آلياً أهميتها من أبعاد متعددة تتجاوز البعد التقني الصرف. فعلى المستوى الحضاري، تعد اللغة العربية من أعرق اللغات الحية، وهي تحمل خصائص تركيبية ودلالية فريدة تجعل من تراثها حقلاً خصباً للبحث اللساني المقارن. إن إتاحة هذا التراث رقمياً يعني "دمقرطة المعرفة"، وجعل أمهات الكتب والمخطوطات النادرة متاحة للباحثين في مشارق الأرض ومغاربها بضغطة زر، مما يسرع وتيرة الإنتاج العلمي ويحقق التراكم المعرفي المطلوب³ (الفهري، 2006، ص. 88).

وعلى المستوى الاقتصادي، تندرج هذه الجهود ضمن ما يعرف بـ "اقتصاد المعرفة" و"الصناعات اللغوية". فبناء مدونات لغوية تراثية ضخمة (Corpora) ومعالجتها يعد المادة الخام لتدريب خوارزميات الذكاء الاصطناعي، وتطوير محركات بحث ذكية، وتطبيقات للترجمة الآلية، والتحليل الدلالي، مما يخلق فرصاً استثمارية هائلة في سوق البرمجيات العالمية. كما أن الحفاظ على التراث يمثل تعزيزاً للأمن القومي الثقافي، حيث يحمي الذاكرة الجماعية من التزييف أو التلاشي في ظل العولمة الجارفة⁴ (زغواني، 2018، ص. 20).

III - الآليات التقنية لمعالجة النصوص التراثية

² - علي، نبيل. (1988). اللغة العربية والحاسوب. الكويت: تعريب للنشر، ص 112.

³ - الفهري، عبد القادر فاسي. (2006). اللسانيات واللغة العربية: نماذج تركيبية ودلالية. بيروت: دار توبقال للنشر، ص 88.

⁴ - زغواني، وجدي. (2018). مدخل إلى المعالجة الآلية للغة العربية. الدوحة: المركز العربي للأبحاث ودراسة السياسات، ص 20.

إن الرحلة من "المخطوطة الورقية" إلى "البيانات المهيكلية" تمر عبر سلسلة معقدة من العمليات التقنية التي تتضافر فيها جهود اللغوي والمبرمج. تبدأ هذه الرحلة بتقنية "التعرف الضوئي على الحروف (OCR)"، وهي العتبة الأولى والأكثر وعورة. فبينما حققت هذه التقنية نجاحاً باهراً مع اللغات اللاتينية والنصوص المطبوعة، فإنها تواجه تحديات جمة مع المخطوطات العربية المكتوبة بخط اليد. ويرجع ذلك إلى طبيعة الخط العربي المتصل (Cursive)، وتعدد أنماط الخطوط التراثية (كالنسخ، والثلث، والمغربي، والكوفي)، وتداخل الحروف، فضلاً عن وجود الهوامش والتعليقات الجانبية (الطرفة) التي تربك الخوارزميات التقليدية⁵ (الحاج، 2015، ص. 30). ولحل هذه المعضلة، تتجه الأبحاث الحديثة نحو استخدام "الشبكات العصبية العميقة (Deep Neural Networks)" التي تحاكي طريقة عمل الدماغ البشري في التعرف على الأنماط، مما مكن من تطوير أنظمة للتعرف على الكتابة اليدوية (HTR) بدقة متزايدة.

بعد تحويل الصورة إلى نص، تأتي مرحلة "التحليل الصرفي الآلي (Morphological Analysis)" فاللغة العربية لغة اشتقاقية بامتياز، تعتمد على نظام الجذر والوزن، وتتميز بترائها الصرفي العالي. فالكلمة الواحدة في التراث قد تكون جملة كاملة (مثل: "فأسقيناكموه")، مما يتطلب من النظام الآلي قدرة فائقة على "التقطيع (Tokenization)" وفصل السوابق واللاحق لاستخراج الجذر الأصلي⁶ (Habash, 2010, p. 65). هذا التحليل هو حجر الزاوية في بناء الفهارس والمعاجم التاريخية، إذ لا يمكن تتبع تطور دلالة لفظة ما عبر القرون دون القدرة على ردها إلى جذرها بدقة.

ولا تكتمل المعالجة إلا بـ "التشكيل الآلي (Automatic Diacritization)" فمعظم النصوص التراثية غير مشكولة، وغياب الحركات في العربية يولد لبساً دلاليًا (Ambiguity) كبيراً؛ فكلمة (عقد) قد تُقرأ (عَقْد، عِقْد، عَقَد). هنا تتدخل الخوارزميات الإحصائية وقواعد النحو الحاسوبي للتنبؤ بالحركة الصحيحة بناءً على السياق، مما يسهل على الآلة وعلينا فهم النص بدقة⁷ (شرف الدين، 2007، ص. 150). وأخيراً، تأتي مرحلة "استخراج الكيانات المسماة (NER)"، وهي تعليم الحاسوب كيفية تمييز أسماء الأعلام، والأماكن، والكتب، والفرق الكلامية داخل النصوص، مما يتيح بناء خرائط معرفية وعلاقات شبكية تربط بين علماء العصور المختلفة ونتائجهم.

VI - التحديات والعقبات: بين صخرة اللغة ومطرقة التقنية:

⁵ - الحاج، حمدي. (2015)، المعالجة الصرفية الآلية للغة العربية: الأسس والتطبيقات. عمان: دار اليازوري العلمية، ص 30.

⁶ - Habash, Nizar Y. (2010). *Introduction to Arabic Natural Language Processing*. Morgan & Claypool Publishers, p. 65.

⁷ - شرف الدين، أحمد. (2007)، الذكاء الاصطناعي ومعالجة اللغة العربية. القاهرة: عالم الكتب، ص 150.

على الرغم من الآفاق الواعدة، إلا أن الطريق نحو رقمنة شاملة وذكية للتراث اللغوي العربي لا يزال محفوفاً بالعقبات. يأتي في مقدمة هذه التحديات "الخصوصية اللغوية للعربية". فالنظام الكتابي واللغوي للعربية يتسم بخصائص شديدة التعقيد مقارنة باللغات الهندية-الأوروبية، مثل ظاهرة "المشترك اللفظي" و"الترادف"، وتغير معاني الكلمات بتغير سياقاتها التاريخية، مما يجعل مهمة الآلة في "الفهم الدلالي" (Semantic Understanding) غاية في الصعوبة⁸ (Farghaly, 2009, p. 44).

يضاف إلى ذلك تحدي "نقص البيانات المرجعية" (Lack of Language Resources) "فالدكاء الاصطناعي لا يتعلم من الفراغ، بل يحتاج إلى "مدونات معيارية" (Gold Standard Corpora)؛ وهي نصوص تراثية تم تحقيقها وتشكيلها وتوسيمها يدوياً بدقة عالية لتكون مرجعاً للتدريب. وللأسف، فإن المحتوى العربي المرقم بدقة لا يزال شحيحاً مقارنة بحجم التراث الفعلي، وما هو متاح غالباً ما يكون بصيغة (PDF) غير قابلة للمعالجة، أو نصوصاً مليئة بالأخطاء الطباعية الناتجة عن النقل اليدوي السريع⁹ (علي، 2001، ص. 210).

كما لا يمكن إغفال التحدي البشري المتمثل في "الفجوة المعرفية" بين المتخصصين. فمهندسو الحاسوب غالباً ما يفتقرون إلى العمق اللغوي لفهم دقائق التراث، في حين يفتقر اللغويون إلى الدراية بحدود الآلة وإمكاناتها. إن غياب العمل المؤسسي المشترك والفرق البحثية "البينية" (Interdisciplinary Teams) "يؤدي إلى جهود مبعثرة ومشاريع فردية تموت بانتهاء تمويلها أو حماس أصحابها"¹⁰ (Soudi & Van den Bosch, 2015, p. 12).

خاتمة :

في ختام هذا الطرح التحليلي المستفيض، الذي جُسنا فيه خلال واقع التراث اللغوي العربي وتحديات دمجها في البيئة الرقمية المعاصرة، نخلص إلى قناعة راسخة ومبدئية؛ مفادها أن قضية الرقمنة في سياقنا الحضاري الراهن لم تعد ترفاً تقنياً، أو مجرد إجراء إداري يهدف إلى أرشفة الماضي وتخزين المخطوطات في خوادم صامتة. بل هي في جوهرها "مشروع أمن قومي معرفي" بامتياز، وورشة حضارية كبرى مفتوحة على آفاق المستقبل، تهدف إلى

⁸ - Farghaly, Ali (Ed.). (2009). *Arabic Computational Linguistics: Current Implementations*. CSLI Publications, p. 44.

⁹ - علي، نبيل. (2001). *الثقافة العربية وعصر المعلومات*. الكويت: سلسلة عالم المعرفة (عدد 265)، ص 210.

¹⁰ - Soudi, A., & Van den Bosch, A. (2015). *Arabic NLP: Overview, State of the Art, Challenges and Opportunities*. In: *Arabic Computational Linguistics*. Springer, p. 12.

استنهاض الذاكرة العربية من حيز "الجمود المادي" والتقدير المتحفّي، إلى حيز "الديناميكية الرقمية" والاستثمار الاقتصادي والفكري. لقد أثبتت فصول هذه الدراسة أن الرهان الحقيقي الذي يواجه الأمة اليوم لا يكمن في مراكمة ملايين "الصور الرقمية (Digital Images)" التي لا تفهم الآلة محتواها، بل في تطوير خوارزميات اللسانيات الحاسوبية والذكاء الاصطناعي لتحويل هذا الإرث الضخم إلى "بيانات ذكية ومهيكلّة (Smart Structured Data)" قابلة للتداول والتحليل والاستنباط في فضاء اقتصاد المعرفة العالمي.

I - النتائج المتوصل إليها:

لقد تبلورت من خلال هذا المسار البحثي مجموعة من النتائج المحورية التي تشكل خارطة طريق للمستقبل؛ أهمها أن التحول الجذري من "الرقمنة التصويرية" التقليدية إلى مرحلة "المعالجة الآلية الذكية" هو السبيل الوجودي الوحيد لكسر طوق العزلة التقنية التي يعيشها التراث العربي، وضمان حضوره النديّ في عصر الذكاء الاصطناعي التوليدي. كما كشف البحث بوضوح عن وجود "فجوة معرفية وتنسيقية" عميقة بين المختصين في علوم اللغة والتحقيق من جهة، والخبراء في هندسة الحوسبة والبرمجيات من جهة أخرى، مما أدى إلى بعثرة الجهود وضياح الموارد في مشاريع فردية تفتقر إلى المعايير الموحدة. وأكدت النتائج أن غياب "الذخائر اللغوية المعيارية (Gold Standard Corpora)" الموثقة والموسومة بدقة يمثل الحلقة المفقودة في السلسلة التقنية؛ فهي "الوقود الحيوي" الذي بدونها تظل أقوى نماذج الذكاء الاصطناعي عاجزة عن فهم نسيج النصوص التراثية أو فك شفراتها الدلالية المعقدة.

II - التوصيات الإجرائية والاستراتيجية:

بناءً على ما تقدم من نتائج، تتقدم هذه الورقة بحزمة من التوصيات الاستراتيجية التي نأمل أن تجد طريقها إلى حيز التنفيذ:

على المستوى المؤسسي: ضرورة التأسيس الفوري لـ "هيئة وطنية أو قومية عليا لرقمنة التراث والذكاء الاصطناعي"، تكون بمثابة مظلة جامعة تتولى وضع استراتيجية موحدة، وصياغة "بروتوكولات معيارية (Standard Protocols)" للفهرسة والترميز (Metadata) وتوحيد صيغ البيانات، بما يضمن تكامل المشاريع ومنع الهدر الناتج عن تكرار الجهود.

على المستوى الأكاديمي والتعليمي: الدعوة إلى إحداث ثورة جذرية في المناهج الجامعية لردم الهوة السحيقة بين العلوم الإنسانية والعلوم البحتة، وذلك من خلال دعم وإنشاء تخصصات "الإنسانيات الرقمية (Digital Humanities)" والتي لا تقتصر على إنتاج جيل من الباحثين "المهجين" الذين يمتلكون ناصية اللغة العربية وفنونها بقدر ما يمتلكون مهارات البرمجة والتعامل مع الخوارزميات.

على المستوى الاقتصادي والتقني: تحفيز القطاع الخاص والشركات الناشئة (Startups) للاستثمار في "التنقيب في البيانات التراثية"، وتحويل هذا الإرث من عبء مادي يتطلب ميزانيات للصيانة والحفظ، إلى مورد اقتصادي مستدام يدر الدخل عبر تطبيقات ذكية في مجالات التعليم، والسياحة الثقافية، والترجمة الآلية المتخصصة، وصناعة المحتوى الرقمي الموثوق.

III-آفاق البحث المستقبلي:

أما عن الآفاق التي تفتحها هذه الدراسة، فإننا نرى أن البحث في هذا المجال لا يزال في بداياته الواعدة. إن الأفق القادم يجب أن يتجاوز مجرد "التعرف على الحروف (OCR) إلى مرحلة" المعالجة الدلالية العميقة "و**"فقه اللغة الحاسوبي**"، حيث يمكن توظيف "النماذج اللغوية الكبيرة (LLMs) "لفهم السياقات التاريخية المتغيرة للألفاظ عبر القرون. كما نطمح إلى رؤية مشاريع بحثية تركز على بناء " خرائط معرفية تفاعلية (Knowledge Graphs) تربط بيننتاجات العلماء وأفكارهم عبر العصور بشكل شبكي، بالإضافة إلى استغلال تقنيات "الواقع المعزز (AR) "و"الواقع الافتراضي (VR) "لاستنطاق المخطوطات وجعلها مادة حية قابلة للتفاعل البشري الرقمي.

ختاماً، إن التراث العربي الرابض تحت غبار القرون في أقبية المكتبات، لا ينتظر اليوم من يمسح عنه الغبار بفرشاة المنقب التقليدي فحسب، بل ينتظر "خوارزميات ذكية" تعيد هندسته، وتفك مغاليقه، وتمنحه صك "الخلود الرقمي" العابر للزمان والمكان. إنها دعوة صادقة لإعادة كتابة تاريخنا بلغة "الصفر والواحد"، ليكون تراثنا جسراً متيناً نعبر به نحو المستقبل، لا قيئاً يشدنا إلى الماضي، وبذلك نضمن أن تظل الهوية العربية حاضرة وفاعلة في وجدان المواطن الرقمي الجديد.

قائمة المصادر والمراجع:

- عبد الرحمن، فوج. (2010)الرقمنة في المكتبات ومراكز المعلومات العربية. القاهرة: دار الفكر العربي، ص 45.
- علي، نبيل. (1988). اللغة العربية والحاسوب. الكويت: تعريب للنشر، ص 112.
- - الفهري، عبد القادر فاسي. (2006). اللسانيات واللغة العربية: نماذج تركيبية ودلالية. بيروت: دار توبقال للنشر، ص 88
- زغواني، وجدي. (2018). مدخل إلى المعالجة الآلية للغة العربية. الدوحة: المركز العربي للأبحاث ودراسة السياسات، ص 20.

- الحاج، حمدي. (2015)، المعالجة الصرفية الآلية للغة العربية: الأسس والتطبيقات. عمان: دار اليازوري العلمية، ص 30.
- شرف الدين، أحمد. (2007)، الذكاء الاصطناعي ومعالجة اللغة العرب ، عالم الكتب، القاهرة .
- علي، نبيل. (2001)، الثقافة العربية وعصر المعلومات . سلسلة عالم المعرفة (عدد 265)
- Habash, Nizar Y. (2010). *Introduction to Arabic Natural Language Processing*. Morgan & Claypool Publishers, p. 65.
- Farghaly, Ali (Ed.). (2009). *Arabic Computational Linguistics: Current Implementations*. CSLI Publications, p. 44.
- Soudi, A., & Van den Bosch, A. (2015). *Arabic NLP: Overview, State of the Art, Challenges and Opportunities*. In: Arabic Computational Linguistics. Springer, p. 12.