اللغة وعالم الرقمنة

مراجعة مقال

م.د. ايمان اسماعيل عايز

amal9988@yahoo.com

الملخص:

إنَّ استخدام اللغة العربية في مجال الرقمنة صار معلوما بالضرورة وان لم نسهم فيه بشكل مباشر وبالقدر المطلوب. كما اننا ندرك ما يشوب من استخدامه من فوضى لا تراعي فيه أصول العربية وقواعدها ناهيك عن ضألة العملية الإنتاجية للغة العربية في هذا الفضاء التكنولوجي ، على الرغم من الجهود التي تبذل في مؤسسات الدول العربية المختلفة بكميات رقمية تتفاوت نسبها بحسب الإمكانات المادية والبشرية والعلمية. و الاهتمام بالمعالجة الآلية للغة العربية ، وتعريب نظم التشغيل ، وتصميم لغات برمجة عربية ، والاستعداد للدخول إلى عصر الترجمة الآلية ، بالرغم من ان هناك جهود في معالجة اللغة العربية آلياً ، مثل الصرف الآلي ، والإعراب الآلي ، والتشكيل التلقائي ، وبناء قواعد البيانات المعجمية ، فالعربية لغوياً وحاسوبياً ، يمكن النظر إليها بلغة الرياضيات الحديثة على أنها لغة عريقة تندرج في إطارها كثير من اللغات الأخرى .

الكلمات المفتاحية: (اللغة، عالم الرقمنة، الترجمة الآلية).

Language and the world of digitalization

Review an article

Dr. Iman Ismail eayiz

amal9988@yahoo.com

Abstract:

The use of the Arabic language in the field of digitization has necessarily become known, even if we do not contribute to it directly and to the required extent. We are also

aware of the chaos that afflicts its use, in which it does not take into account the origins and rules of Arabic, not to mention the insignificant production process of the Arabic language in this technological space, despite the efforts being made in the institutions of various Arab countries in digital quantities whose percentages vary according to the material, human and scientific capabilities. And interest in the automatic processing of the Arabic language, the Arabization of operating systems, the design of Arabic programming languages, and the preparation to enter the era of machine translation, although there are efforts in processing the Arabic language automatically, such as automatic morphology, automatic parsing, automatic formation, and building lexical databases. Linguistically and computationally, it can be viewed in the language of modern mathematics as an ancient language within which many other languages fall.

Keywords: (language, digital world, machine translation).

مقدمة:

إنَّ الحاجة إلى حفظ الثقافة والتقاليد سابقا وحث ابناء المجتمع الواحد على ذلك دفعا لتعلم القراءة والكتابة لرجال الكهوف. وأصبحت العملية التدريجية لعمل العلامات والكتابة الأولية مصدر إلهام قاده إلى هذا الفن المتمثل في إعطاء شكل مادي للكلمات التي يتحدثون بها. من خلال التقليد الذي شوهد على ألواح الطين والبردي والورق والآن يجد مكانه في قاعدة البيانات. وعلى الرغم من أنها تبدو أحيانًا بطيئة ، إلا أنها شهدت تحولًا جذريًا في فترة قصيرة من ألفي عام. رقمنة اللغة ليست بالأمر الجديد بل هي طريقة لتوثيق اللغة لحفظ البيانات الموجودة في أشكال الكتاب والعديد من المخطوطات. لا شك أنها لم تأخذ تلك النظرة الأوسع .

فلقد استحدثت التقنيات في الاونة الاخيرة من القرن العشرين من كرات الالة الكاتبة الى الهواتف المحمولة المبكرة واليوم اصبحت رقمية بالكامل من انظمة التشغيل وتطبيقات سطح المكتب الى الرسائل النصية وتطبيقات الاجهزة المحمولة مثل الخرائط ووسائل التواصل الاجتماعي وكل شيء اخر تكنولوجي يتواصل به ابناء البشر الواحد .

حيث ان هناك أكثر من ٩٩٪ من المحتوى على الإنترنت مكتوب بواحدة من ٣٥ لغة مختلفة. حيث يتم إرسال ٦٠٠٠ لغة متبقية أو نحو ذلك إلى زوايا صغيرة نسبيًا على الويب ، مما يجعل الرقمنة مفتاحًا للحفاظ على اللغة حيث أصبح الإنترنت متأصل بشكل متزايد في المجتمع العالمي. مع جولة كبيرة من المنح المقدمة من (National Endowment for the Humanities) التي تتجه نحو مشاريع رقمنة اللغة الأمريكية الأصلية في وقت ليس بالقليل ، جنبًا إلى جنب مع العديد من المشاريع الحديثة الأخرى لرقمنة اللغات الأصلية في جميع

أنحاء العالم ، مما يجدر استكشاف كيف يمكن لهذه المشاريع أن تساعد في تنشيط اللغة.

و نجد ان اللغة الإنجليزية هي اللغة الأكثر شيوعًا على الإنترنت ، وتشكل ما يزيد قليلاً عن ٦٢٪ من المحتوى المكتوب عبر الإنترنت – تليها الروسية والإسبانية بنسبة ٥,٩٪ و ٣,٧٪ على التوالي. هذه الأرقام ليست بأي حال من الأحوال وصفًا دقيقًا لمدى انتشار كل لغة على نطاق واسع ، حيث يتم التحدث باللغة الإنجليزية من قبل ما يقرب من ١٨ ٪ من العالم بأسره ، في حين أن لغة الماندرين الصينية ، التي يتحدث بها حوالي ١٤ ٪ من سكان العالم ، لا تشكل سوى ١.٤ ٪ من محتوى الإنترنت .

هذا ويؤدي التمثيل المفرط للغة الإنجليزية (وبعض اللغات الأخرى مثل الفارسية ، على سبيل المثال) إلى نقص تمثيل اللغات الأخرى ، مع وجود لغات أقل تحدثًا وقليلة الموارد تتحمل العبء الأكبر من هذا التأثير كاللغة العربية . نظرًا لأن الإنترنت يلعب دورًا

كبيرًا في عادات استهلاك الأشخاص لوسائل الإعلام ، يمكن أن يؤثر وجود اللغة على الإنترنت – أو عدم وجودها – على طول عمرها ، لا سيما في الحالات التي تكون فيها اللغة معرضة بالفعل للخطر.

وقبيل التبني الواسع للإنترنت – وأجهزة الكمبيوتر ككل ، حقًا – كان إحياء اللغة أكثر صعوبة مما هو عليه اليوم (والذي ينبغي أن يقول الكثير مع الأخذ في الاعتبار حقيقة أنه ليس سهلاً اليوم أيضًا). في جهود إحياء اللغة ما قبل الرقمية ، غالبًا ما كان على اللغويين قضاء قدر كبير من الوقت في تطوير القواميس المطبوعة والقواعد النحوية بناءً على التفاعلات مع المتحدثين الأصليين المتبقين .

ولا يزال يتعين على اللغويين بذل جهود مماثلة تستغرق وقتًا طويلاً ، ولكن تخزين البيانات اللغوية عبر الإنترنت يوفر مساحة فعلية ويخلق موارد يسهل الوصول إليها. هذا يجعل الأمور أكثر سهولة للباحثين ومجتمعات اللغات على حد سواء لمشاركة المعلومات حول اللغة. علاوة على ذلك ، يساهم نشر المحتوى المكتوب والمنطوق عبر الإنترنت بلغة معينة في مجموعة المعلومات المتاحة للجمهور في اللغة – وقد استفادت مجموعات مثل (WikiAfrica) من ذلك للمساعدة في نشر المعلومات بلغات أفريقية مختلفة لا يتم استخدامها على نطاق واسع عبر الإنترنت.

لذا نلاحظ ان اللغة العربية تمر بوضع تاريخي حديث في عصر الرقمنة الحالية مما أدى بها الى استخدام معالجة فعلية واقعية لاكتساب المعرفة أمام تحديات متنوعة الأوجه . وتتجسد هذه القوة في تو سيع التكنولوجية المتطورة التي يزيد تعميمها يوميا في العديد من مجالات الحياة خصوصا الرقمية التي تكاد تطغى على ممارسة حياة الافراد ومعاملات المؤسسات العمومية والخاصة. والاهتمام بالهندسة المعلوماتية من اجل تطويع اللغة العربية للحوسبة وان كانت هذه المحاولات التكنولوجية لم ترقى الى مستوى الحوسبة

الهندسية في العالم .بسبب عدد من المعوقات على الرغم مما تملكه من قدرات وادوات معلوماتية . لذا تستحق اللغة العربية من مجتمعاتها الجهود المضنية والحثيثة لتطوير رقمنتها من خلال إثراء المحتوى العربي على الإنترنت ، ومن خلال عدد من البرمجيات الذي يأخذ على عاتقه دعم هذا المحتوى العربي، إلى جانب العديد من المشاريع الاستراتيجية المهمة والمبادرات التي تعمل رقمنة محتواها ، ، ومن أهم هذه المبادرات والمشاريع ما يلي :

مشروع ذاكرة العالم الذي أطلقته منظمة الأمم المتحدة للتربية والثقافة والعلوم (اليونسكو) للحفاظ على اللغة العربية فالتراث العربي كنز نادر يجب صيانته والحفاظ عليه ،من أجل الحفاظ على الثروة الثقافية العربية . ويعد هذا المشروع واحداً من المشروعات الاستراتيجية العربية للاتصالات والمعلومات ، وتنفذه وزارة الاتصالات وتكنولوجيا المعلومات بمصر الشقيقة ، متمثلة بمركز توثيق التراث الحضاري والطبيعي وبمساهمة منظمة اليونسكو ، والاتحاد الدولي للاتصالات ، ومكتبة الإسكندرية ، ومن اهداف هذا المشروع : رقمنة المحاور الحضارية المختلفة و الحفاظ على الذاكرة التراثية الإجمالية لشعوب العالم العربي ، وتعريف الأجيال الجديدة بها و تشجيع أعمال الرقمنة في الدول العربية ، والمساهمة في زيادة المحتوى الإلكتروني العربي ونشره على الإنترنت و البناء والاستفادة مما تم انجازه من مبادرات عربية وإقليمية في مجالات توثيق التراث والثقافة العربية .

ولكي يتم دعم اللغة على الإنترنت ، يجب أن تحدث سلسلة من الخطوات : وهي إنشاء كومة من المستندات حول كيفية الرقمنة. وتعني رقمنة اللغة ترميز النص الخاص باللغة بحيث يمكن التعرف عليه رقميًا وإنشاء برنامج للوحة المفاتيح الخاصة بها ووسائل الإدخال الأخرى وإنشاء قوالب لها لبدء تجميع المحفوظات الرقمية للصوت والنص ،

وإنشاء قواميس رقمية ودورات رقمية للأطفال لتعلم لغتهم من خلال الأدوات الحديثة. و الوصول إلى المجتمعات وتقييم احتياجاتها الرقمية ، و توجيه المجتمعات من خلال تنفيذ رقمنة اللغة.