

د. مصطفى جرّار يستعرض جهود جامعة بيرزيت في حوسبة اللغة العربية

مجتمع الجامعة - قصص مميزة - أبحاث ومشاريع - علوم وتكنولوجيا

25 شباط 2019



استعرض أستاذ علم الحاسوب (<https://www.birzeit.edu/ar/content/lm-lhswb>) في جامعة بيرزيت د. مصطفى جرّار (ar/faculty-staff/mstf-jrr/) الجهود التي بذلها هو وفريق عمله على مدار أكثر من ثماني سنوات من أجل حوسبة اللغة العربية، وإثراء المحتوى الرقمي باللغة العربية عبر شبكة الانترنت، وذلك من خلال: محرك البحث المعجمي، ومحرك "الانطولوجيا العربية" (<https://ontology.birzeit.edu>)، ومحرك "كزّاس" للهِجة العاميّة الفلسطينية.. والمتاحة جميعها لاستخدام الباحثين والجمهور مجاناً عبر موقع الجامعة الإلكتروني.

هذا العرض جاء ضمن ندوة "الانطولوجيا العربية وحوسبة المعاجم" التي نظّمها منتدى النهضة العربية - مركز النهضة الفكري في الأردن، والتابع لمنظمة النهضة العربية للديمقراطية والتنمية (أرض)، وذلك بالتزامن مع "اليوم العالمي للغة الأم" (<https://www.birzeit.edu/ar/18-knwn-lwl-ywm-lly-llg-lrby>) الذي تحتفل به منظمة "اليونسكو" في (21) شباط/ فبراير من كل عام.

ويعتبر "محرك البحث المعجمي" (<https://ontology.birzeit.edu>) لجامعة بيرزيت الأول من نوعه في العالم ليس فقط على مستوى اللغة العربية، بل على مستوى جميع اللغات المستخدمة عبر شبكة الانترنت، حيث تتضمن قاعدة بيانات المحرك حوسبة (150) معجماً عربياً مختلفاً، سواء المعاجم اللغوية العامّة التراثية والحديثة، أو المعاجم العلمية المتخصصة في مختلف مجالات العلوم التطبيقية والإنسانية.

ويؤقر المحرك ترجمة دقيقة بالعربية للمصطلحات والألفاظ الأجنبية، وبما يتجاوز الأخطاء والمغالطات الكثيرة التي تعتري ترجمات محركات البحث شائعة الاستخدام مثل محرك (جوجل ترانزليت)، حيث تعتمد المحركات على غرار (جوجل) ما يسمى "الترجمة الإحصائية" لمعاني المفردات والكلمات، وليس الترجمة "المعجمية" المنضبطة كما في محرك جامعة "بيرزيت".

واستدعى العمل على حوسبة المعاجم العربية إعادة طباعة محتواها يدويا من أجل الحصول على نسخة إلكترونية من النص (text) لغايات برمجتها وإدخالها ضمن قاعدة بيانات محرك البحث، حيث تم الاستعانة في مرحلة من المراحل بجهود أكثر من (300) طالبا من طلاب الجامعة للمساعدة في هذه العملية.

وزاد من صعوبة العمل اعتماد جامعة "بيرزيت" وفريق العمل على جهودهم وإمكاناتهم الذاتية في إنجاز المشروع، من دون الحصول على أي دعم مادي من أي جهة كانت من أجل تمويل مشروع بهذا الحجم، وهو ما ينطبق أيضا على محركي "الانطولوجيا العربية" و"كّراس".

ويهدف محرك "الانطولوجيا العربية" إلى ضبط فوضى التعريفات والدلالات والمفاهيم في اللغة العربية، والتي تعتري حتى النقاشات العلمية والأكاديمية المتخصصة، حيث تُستخدم الكلمة الواحدة أو المصطلح الواحد من قبل أكثر من طرف ضمن النقاش الواحد أو السياق الواحد، في الوقت الذي يكون لكل طرف من هذه الأطراف معناه الخاص ودلالته الخاصة لهذا المصطلح، الأمر الذي يقود إلى سوء الفهم والالتباس والخلاف.

ويستعير محرك "الانطولوجيا العربية" مفهوم "الانطولوجيا" من الفلسفة، والذي يعني "علم الوجود"، أو علم "ماهية الأشياء"، أو "علم الوجود بما هو موجود" كما تشير إليه المصادر التراثية العربية.. ويزاوج بين هذا المفهوم الفلسفي وعلم اللغة وعلم الحاسوب، بكون مثل هذه المزاجية هي الأساس المنطقي الذي تقوم عليه تطبيقات ما يسمى "الذكاء الصناعي".

وقد تم تصميم محرك "الانطولوجيا العربية" ليتناسب بشكل أساسي مع استخدامات أجهزة الحاسوب، وبطريقة يمكن تطويرها والبناء عليها في المستقبل للحصول على تطبيقات وأجهزة ذكية (كالروبوتات) تستطيع التعرف على اللغة العربية صوتيا، وفهم معني الكلام الموجه إليها، والرد عليه تلقائيا الرد المناسب (التعلم الذاتي).

أما بالنسبة لمحرك "كّراس" فهو يتيح لمحركات البحث وأجهزة الحاسوب التعرف على المحتوى المكتوب والصوتي عبر شبكة الانترنت باللغات العامية، إضافة إلى توثيق هذه اللهجات، حيث تم الابتداء بحوسبة بعض اللهجات العامية الفلسطينية الدارجة.

ويأتي محرك "كّراس" استجابة لحقيقة واقعية مفادها أن نسبة كبيرة من المحتوى العربي عبر شبكة الانترنت مكتوب باللغات العامية، وعدم قدرة محركات البحث وأجهزة الحاسوب على التعرف على هذه اللهجات يعني ببساطة أن يبقى هذه الكم الكبير من "المعرفة العربية" و"الخبرة العربية" بعيدا عن متناول الباحثين والدارسين والحصيلة المعرفية الإنسانية عبر الشبكة.

وباستثناء هذه الحاجة التقنية، لا يهدف محرك "كّراس" للترويج للهجات العامية وإحلالها مكان اللغة الفصحى، بل على العكس، يقوم بضبط مفردات وكلمات هذه اللهجات العامية المختلفة من خلال إيراد معانيها ومترادفاتها الدلالية باللغة العربية الفصحى.

واختتم "جرار" عرضه بالدعوة لإرساء "صناعة لغوية عربية" متكاملة في العالم العربي، سواء لوجود فرصة سانحة وسوقاً واعدة كبيرة في هذا المجال حالياً، أو لكون هذا حاجة أساسية تتعلق بالأمن "الوجودي" للغة العربية ثقافياً واجتماعياً بما يتجاوز الاستخدامات التطبيقية والوظيفية للغة.

من جانبهم تفاعل الحضور مع العرض الذي قدّمه الدكتور "جرار" من خلال أسئلتهم التي توزعت بين أسئلة تتعلق بالمفهوم الفلسفي للانطولوجيا نفسها، وأسئلة تتعلق بواقع اللغة العربية ومستقبل تدريسها واستخدامها بشكل عام، وأسئلة تتعلق بالجوانب الفنية والتطبيقية لمحركات البحث الثلاثة.

كما تميّزت الندوة، بما ينسجم مع منحها العلمي المتخصص، بمدخلات نوعية قام بها أكاديميون وطلبة جامعيون متخصصون في مجال هندسة الحاسوب والبرمجيات، ومدخلات لخبراء سبق لهم العمل أو يعكفون حالياً على مشاريع ومنتجات تتعلق بحوسبة اللغة العربية واستخداماتها الالكترونية وتطبيقاتها الذكية عبر الشبكة.

وكان رئيس مجلس أمناء منظمة النهضة العربية (أرض) الدكتور زيد عيادات قد تولّى إدارة الندوة والأسئلة والمدخلات التي أعقبها، حيث بيّن في معرض تقديمه أن اختيار (أرض) لموضوع "الانطولوجيا العربية وحوسبة المعاجم" يعود إلى كون اللغة و"الأمن اللغوي" هما من المقومات الأساسية التي يقوم عليها المشروع النهضوي الشامل الذي تسعى منظمة النهضة العربية لإرسائه.

يُذكر أن الدكتور مصطفى جزار حاصل على شهادة الدكتوراه من جامعة بروكسل الحرّة/ بلجيكا سنة (2005)، وسبق له الحصول على منحة الزمالة (ماري كوري) بين عامي (2007 – 2009)، وزمالة (فلبرايت) سنة (2017) عمل خلالها أستاذاً زائراً في كلية الطب بجامعة نيويورك بالولايات المتحدة الأمريكية. ولديه أكثر من سبعين بحثاً علمياً محكماً في مجالات: الذكاء الصناعي، وحوسبة اللغات، وهندسة البيانات، والحوكمة الإلكترونية. وهو مؤسس "معهد ابن سينا لهندسة المعرفة والتقنيات العربية" بجامعة "بيرزيت"، ومؤسس "أكاديمية فلسطين للحوكمة الإلكترونية".