وحدة دراسة المجال الاجتماعي الرقمي العربي
المركز العربي للأبحاث ودراسة السياسات، الدوحة، قطر

Unit for Research in Arab Digital Social Spaces
Arab Center for Research and Policy Studies, Doha, Qatar

الرقم التعريفي DOI
https://doi.org/10.31430/TJOF9352

مقدمة

شهد العقدان الأخيران تطورًا هائلاً في حجم البيانات والمحتوى المنشور على الإنترنت عالميًّا وعربيًّا. وأتاحت وسائط التواصل الاجتماعي ميلاد نمط جديد من المحتوى الرقمي، يتميز بتعدد مصادره وانخفاض تكلفة إنتاجه، إضافةً إلى تنوع موضوعاته وصوره. وبات الأفراد العاديون مساهمين فاعلين في إنتاج المحتوى على نحو غير مسبوق، سواء كان ذلك من خلال تفاعلهم مع الموضوعات القائمة أو عبر التعبير عن آرائهم في مختلف القضايا والموضوعات. ويميز هذا التحوّل الجيل الثاني من الشبكة العنكبوتية (Web 2.0).

تطور مفهوم شبكة الجيل الثاني ليشير بطريقة أوسع إلى البرمجيّات الاجتماعية (Social Software) التي تسهّل للمستخدمين إنتاج البيانات والتفاعل معها ومع المستخدمين الآخرين. وتزامنت هذه التطورات مع ظهور مفهوم البيانات الكبيرة (Big Data) التي أصبحت ممكنةً مع ارتفاع سعة التخزين، وسرعة إنتاج البيانات وتحليلها ومعالجتها، وتدنّي تكلفتها أيضًا. وأخذ تطوّر علم البيانات، بوصفه حقلاً أكاديميًّا جديدًا، منحًى عابرًا للتخصصات تبعًا لطبيعة البيانات المتعددة الأنواع والمحتوى، بحيث أصبح يدمج علوم الحوسبة والإحصاء في مجالات كانت بمنأى عنها مثل الاقتصاد، والقانون، والتاريخ، والأدب، والاجتماع، والسياسة، والتاريخ.

وفي هذا السياق، ظهرت مختبرات ووحدات بحثية متخصصة في تحليل البيانات[1]، انبثقت من جامعات ومراكز بحثية في مختلف أنحاء العالم. وبعد أن كان مجال التزاوج بين علوم الحوسبة والعلوم الإنسانية والاجتماعية ينحصر في "العلوم الإدراكية" (Cognitive Sciences)، و"اللسانيات الحاسوبية" (Computational Linguistics)، انطلق مع تقنيات "معالجة اللغات الطبيعية" (Processing and Natural Language) إلى مجالات أخرى مثل "تحليل الشبكات الاجتماعية" (Social Network Analysis)، و"العلوم الاجتماعية الحاسوبية" (Computational Social Science)، و"الإنسانيات الرقمية" (Digital Humanities).

وتقدّم اللسانيات الحاسوبية ومعالجة اللغات الطبيعية نماذج لتحليل النص المطبوع والمكتوب والصوت والصورة لفهمها واستخراج المعلومات منها وتوليد ما يشبهها. وتستخدم لذلك الخوارزميات الدقيقة، والنماذج الإحصائية التقريبية، ونماذج التلقين الآلي والذكاء الاصطناعي الناتجة من البيانات. ويستفيد تحليل الشبكات الاجتماعية من كل هذه التقنيات لاستخراج الكيانات والعلاقات في شكل شبكات، ويركز على دراسة بنية هذه العلاقات الاجتماعية من خلال نظرية الرسوم البيانية (Graph Theory)، والنماذج الحاسوبية. وبذلك، يقدّم فهمًا لشبكات يشكّلها أفراد ومنظمات وجماعات. ويستعين علم الاجتماع الحاسوبي بمجموعة من الأساليب التقنية

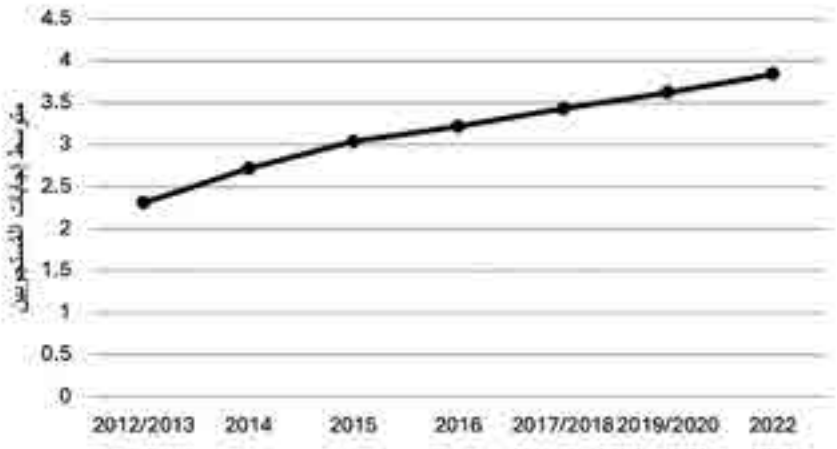
1 منها، على سبيل المثال، معهد العلوم الاجتماعية الكمية بجامعة هارفارد (Institute for Quantitative Social Science, Harvard University)، ومعهد البيانات والأنظمة والمجتمعات بمعهد ماساتشوستس للتكنولوجيا (Institute for Data, Systems, and Societies, MIT)، ومركز كوبنهاغن لعلوم البيانات الاجتماعية (Copenhagen Center for Social Data Science, SODAS)، ومعهد أكسفورد للإنترنت (Oxford Internet Institute).

مثل "التنقيب عن البيانات" (Data Mining)، و"التلقين الآلي" (Machine Learning)، و"تحليل الشبكات" (Network Analysis)، لدراسة الظواهر الاجتماعية. وتمتد هذه المقاربة الرقمية إلى عموم العلوم الإنسانية مثل الأدب، والتاريخ، والفن، والدراسات الثقافية.

وشكّلت انتفاضات الربيع العربي عام 2011 لحظةً مفصلية في تطور العلوم الاجتماعية الحاسوبية وتحليل الشبكات الاجتماعية، تزامنًا مع تطور وسائط التواصل الاجتماعي وتحوّلها إلى منصات حيوية للتفاعل بين الأفراد والمؤسسات و"المؤثرين في الرأي العام". فقد أصبحت هذه الوسائط أداةً قويةً للتعبير السياسي والتعبئة، وهو ما أتاح كمًّا هائلًا من البيانات والمعلومات، وجعلها بيئة خصبة للبحث في الظواهر السياسية والاجتماعية التي لم يكن في الإمكان دراستها من قبل. ومنذ ذلك الحين، شهد الحضور العربي على الإنترنت نموًّا ملحوظًا، كما هو واضح في الشكل (1) الذي يستعرض تقدير المستجيبين الذاتي للوقت الذي يقضونه على الإنترنت، استنادًا إلى بيانات "المؤشر العربي"[2]. ويقيس المتغيرُ معدلات الاستخدام وفقًا لفئات تراوح بين "لا أستخدم الإنترنت" (1)، وصولًا إلى "عدة مرّات في اليوم/ مرة واحدة في اليوم" (5). ومع ذلك، لا تزال معظم الدراسات التي تناولت المحتوى العربي صادرة عن باحثين أجانب؛ ما يؤكد الحاجة إلى مزيد من الجهود البحثية العربية في هذا المجال.

الشكل (1)

متوسط استخدام الإنترنت في العالم العربي



2 هو استطلاع سنوي لقياس اتجاهات الرأي العام العربي في القضايا السياسية والاقتصادية والاجتماعية، أطلقه المركز العربي للأبحاث ودراسة السياسات عام 2011، وأُجريت دورته الثامنة في 2021 / 2022. ينظر: المركز العربي للأبحاث ودراسة السياسات، **المؤشر العربي**، "برنامج قياس الرأي العام"، شوهد في 2025/4/4، في: https://acr.ps/1L9zQyC

أولًا: رؤية وحدة دراسة المجال الاجتماعي الرقمي العربي

تتفرع رؤية وحدة دراسة المجال الرقمي والاجتماعي العربي من رؤية المركز العربي للأبحاث ودراسة السياسات، الذي تأسس عام 2010، والذي يسعى لتعزيز البحث العلمي المنهجي وفهم قضايا المجتمع والدولة، ويتبنى رؤية نهضوية للمجتمعات العربية من خلال البحوث النظرية والتطبيقية، والنهوض باللغة العربية في العلوم الاجتماعية والإنسانية. ويتبنى المركز العربي رؤية منهجية تكاملية تقوم على تداخل الاختصاصات لدراسة القضايا العربية المعقدة، بحيث لا يمكن فهمها في ضوء الأحادية المنهجية، أو تصنيم الاختصاص، مع الإبقاء على أهمية الاختصاص نفسه رافدًا ضمن رؤية منهجية أشمل[3].

وقد سعى المركز العربي جاهدًا، منذ نشأته، لتعزيز المحتوى الأكاديمي والبحثي العربي. ولتحقيق هذا الهدف، لجأ إلى النشر الأكاديمي المحكّم الذي تجاوز 650 كتابًا و200 عدد من الدوريات العلمية المختصّة[4]. وأطلق المركز العربي عددًا من المشاريع البحثية المعرفية الكبرى، على غرار "برنامج قياس الرأي العام" للمؤشّر العربي[5]، و"معجم الدوحة التاريخي للغة العربية"[6]، ومشروع "توثيق الذاكرة الفلسطينية"[7]، ومشروع "الذاكرة السورية"[8]. وبرزت الحاجة إلى استخدام تقنيات حوسبية متقدمة إبان تطوير هذه المشاريع للتعامل مع الكمّ الهائل من البيانات التي لم تعد الطرائق التقليدية قادرة على مواكبتها.

وفي حزيران/ يونيو 2022، أطلق المركز العربي وحدة دراسة المجال الاجتماعي الرقمي العربي، استجابةً للتطورات المتسارعة في مجال البيانات الكبرى، ونظرًا إلى حاجة المجتمع البحثي العربي إلى هذا النوع من الوحدات المتخصصة. وترافق ذلك مع مجموعة من التطورات التقنية، التي يأتي في مقدّمتها النمو الأسّي السريع في المحتوى الإلكتروني العربي على الإنترنت، وفيض المعلومات بشأن المجال الاجتماعي العربي بعد انتفاضات الربيع العربي 2011؛ وهو ما يتعذّر سبرُ أغواره باستخدام الأدوات البحثية التقليدية.

تتطلع الوحدة إلى رفد البحث الأكاديمي العربي من خلال ابتكار برمجيات وأساليب معرفية تستثمر محتوى المجالات الرقمية العربية لفهم أعمق وأكثر تركيبًا للمجتمعات العربية؛ وذلك من خلال عنايتها بجمع المحتوى العربي وتنضيده

3 **المركز العربي للأبحاث ودراسة السياسات: في عشر سنين** (الدوحة/ بيروت: المركز العربي للأبحاث ودراسة السياسات، تشرين الثاني/ نوفمبر 2022)، ص 10-12.

4 ينظر: المركز العربي للأبحاث ودراسة السياسات، شوهد في 2025/4/4، في: https://acr.ps/1L9zQCP

5 ينظر: المركز العربي للأبحاث ودراسة السياسات، **المؤشر العربي**، "برنامج قياس الرأي العام"، شوهد في 2025/4/4، في: https://acr.ps/1L9zRgX

6 ينظر: معجم الدوحة التاريخي للغة العربية، شوهد في 2025/4/4، في: https://acr.ps/1L9zR04

7 ينظر: المركز العربي للأبحاث ودراسة السياسات، **برامج ومشاريع**، توثيق الذاكرة الفلسطينية، شوهد في 2025/4/4، في: https://acr.ps/1L9zR7R

8 ينظر: "الذاكرة السورية"، مشروع علمي بحثي أنشأه المركز العربي للأبحاث ودراسة السياسات، مؤسسة الذاكرة السورية، شوهد في 2025/4/4، في: https://acr.ps/1L9zQPU

في صيغة بيانات كبرى وفقًا للأصول الأكاديمية والأخلاقيات العلمية، ليكون مصدرًا من مصادر التحليل في البحث الأكاديمي. وتتجه الوحدة نحو تطوير أدوات رقمية لتحليل معمَّق لاتجاهات التفاعل عبر وسائط التواصل الاجتماعي في المنطقة العربية، سواء كان ذلك بين الأفراد أو المجموعات أو المؤسسات. وتُعنى، أيضًا، بدراسة المجتمعات العربية وتفاعلاتها في العالم الافتراضي، للوقوف على تفاعلاتها الداخلية والبينية ومدى تأثيرها وتأثّرها.

وترى الوحدة أنّ لها دورًا فاعلًا في تأسيس بناء الموارد والمعارف العربية الرقمية واستكمالها في مختلف المجالات المعرفية، وتطمح إلى أن تكون رائدةً في تخريج كوادر وباحثين متميزين في إنتاج معارف معاصرة تجمع بين الأساليب الحوسبية، والعلوم الاجتماعية والإنسانية، في المجال العربي.

وتسعى الوحدة لتعزيز الاستقلال العربي في مجال التقنيات الرقمية المتقدمة من خلال استثمار الموارد اللغوية والمعرفية العالية الجودة التي يمتلكها المركز العربي، إلى جانب الموارد اللغوية والحوسبية المتوفرة بصيغة مفتوحة المصدر، في بناء أدوات وتقنيات ونماذج لغوية. وتراعي هذه النماذج الخصوصية الثقافية العربية، وتلبّي احتياجات الباحث العربي، وتبرز أهميتها مع التطور المتسارع في مجال الذكاء الاصطناعي التوليدي والتحويلي؛ إذ يشكل احتكار شركات معدودة مثل "أوبن إي آي" (OpenAI)، و"غوغل" (Google)، و"أنتروبيك" (Anthropic)، لهذه التقنيات تحديًا كبيرًا؛ ما يمنحها سلطةً غير مسبوقة في تشكيل المعارف والأدوات المعرفية في مختلف أنحاء العالم، وهو ما يجعل من الضروري إطلاق مبادرات عربية قادرة على تكييف هذه التطورات لمصلحة المنطقة، وحماية سكانها من الاعتماد على الخارج، درءًا لما قد يترتب على ذلك من آثار سلبية في الخصوصية وفرص التنمية الاقتصادية والتقدم الأكاديمي.

ثانيًا: مخرجات وحدة دراسة المجال الاجتماعي الرقمي العربي

تتنوع مخرجات وحدة دراسة المجال الرقمي والاجتماعي العربي بين مشاريع وخدمات تقدّمها للباحثين في العلوم الاجتماعية والإنسانية، إضافة إلى الأبحاث التي تعمل عليها الوحدة، والتي تسعى لمراكمة المعرفة العربية في مجالات العلوم الاجتماعية الحوسبية ومعالجة اللغة الطبيعية.

أما على صعيد الخدمات، فتقدّم الوحدة خدمات استشارة وتدريب للجهات البحثية، أفرادًا كانوا أم مؤسسات، وتشمل هذه الخدمات ما يلي:

- مساعدة الباحثين على تطوير الأسئلة البحثية في ميدان العلوم الاجتماعية والإنسانية، للاستفادة من الإمكانات المعلوماتية والحوسبية، والإجابة عنها وفق منهجيات علمية معاصرة؛ وذلك من خلال تقديم الاستشارات اللازمة لهم حول كيفية صياغة الأسئلة البحثية بطريقة تسمح باستخدام الأدوات الحوسبية في الإجابة عنها، وتوجيههم نحو المنهجيات والأساليب الملائمة لطبيعة أسئلتهم.
- توفير البيانات اللازمة للباحثين من أجل مشاريعهم البحثية من خلال أتمتة عملية

جمع المحتوى العربي ورصده على مواقع التواصل الاجتماعي، ومعالجته ليكون قابلًا للتحليل؛ إذ تطوّر الوحدة الأدوات والبرمجيات اللازمة لتجميع البيانات من المنصات الاجتماعية بطريقة آليّة، ثم تعمل على تنظيف هذه البيانات وهيكلتها وإعدادها بالصيغة المناسبة، ليتمكن الباحثون من تحليلها واستخراج رؤى ونتائج مختلفة منها.

- تقديم ورشات تدريبية لتحسين قدرات الباحثين على استخدام الأدوات والتقنيات الحسابية واستخراج المعلومات والمعطيات والنصوص المتعلقة بمجالات عملهم، وتحليلها. وتشمل هذه الورش التدريب العملي على استخدام لغات البرمجة وحزم البرمجيات الإحصائية والنصية، وأساليب التنقيب في البيانات وتحليلها، وعرض النتائج بطرائق بصرية فعالة، على نحو يمكّن الباحثين من تطبيق هذه المهارات في أبحاثهم.
- توفير استشارات ودعم تقني لمشاريع المركز العربي التي تتعامل مع بيانات كبيرة، مثل مشروع توثيق الذاكرة الفلسطينية والذاكرة السورية. وتقدّم الوحدة خبراتها للمساعدة في إدارة الكم الهائل من البيانات في هذه المشاريع من الناحية التقنية، وتصميم البنية التحتية المعلوماتية اللازمة لمعالجة كميات ضخمة من البيانات وتحليلها، وتطوير بعض الحلول البرمجية لها.

وفي سياق المشاريع البرمجية والبحثية، نتبيّن ما يلي:

- تعمل الوحدة بالتعاون مع برنامج المؤشّر العربي على إصدار تقرير دوري مُعمّق حول المحتوى العربي على مواقع التواصل الاجتماعي من حيث الاتجاهات العامة، وأنماط التفاعل، والقضايا الأكثر تداولًا، فضلًا عن دراسة تحولات الرأي العام العربي. ويعتمد التقرير على منهجيات بحثية متقدمة تجمع بين التحليل الكمي والكيفي، وتوظيف أحدث تقنيات تحليل البيانات الضخمة لرصد التطورات والمستجدات في الخطاب الرقمي العربي؛ ما يساهم في فهم ديناميكيات النقاش العام وتوجهات المستخدمين على هذه المنصات.
- أنجزت الوحدة 14 "مقياس أداء" (Benchmarks) لتقييم أداء النماذج اللغوية الكبيرة، وذلك من حيث المقاييس التقليدية للأداء، إضافة إلى المقاييس الفريدة المتعلقة بالثقافة والتحيز. وتشمل هذه الاختبارات تقييم المعرفة الأساسية (مثل أسئلة الكتب الدراسية)، والاختبارات اللغوية (مثل التعرف إلى الكيانات المسماة والدلالات)، واختبارات الثقافة والتحيز (مثل التحيز في قضية فلسطين وإسرائيل، والخريطة الثقافية). وتتميز هذه المقاييس باعتمادها على بيانات خاصة بالمركز العربي، علاوة على بيانات لم تُدرج مسبقًا في تدريب النماذج اللغوية الكبيرة؛ ما يمنحها قيمةً إضافيةً وموثوقيةً أعلى.
- تعمل الوحدة على تطوير "نموذج لغوي تأسيسي" (Foundation Model)، يدعى "دلّة" (DALLA)، يلائم احتياجات

المستخدم العربي، ويجتنب الانحيازات التي تَسِمُ النماذج الشائعة مثل "تشات جي بي تي" (ChatGPT)، و"جيميناي" (Gemini)؛ وذلك من خلال الاستناد إلى نماذج لغوية مفتوحة المصدر ورفدها ببيانات عربية تعزز أداء النماذج اللغوية وتخفف من تحيّزها الثقافي. وفي هذا السياق، نجحت الوحدة في تطوير حلول لخفض تكاليف تدريب النماذج اللغوية المفتوحة المصدر، بحيث يكون أداؤها موازيًا لأداء النماذج المكلفة والمغلقة المصدر، في المجالات التي تهم المركز العربي والمجتمع الأكاديمي عربيًا[9].

- بنت الوحدة نماذج لغوية متخصصة تساعد على التصنيف الآلي للنصوص العربية[10] (Topic Classification)، وتستخرج منها الكيانات المسماة والعلاقات فيما بينها (Relational Entity Extraction)[11]، وتستخدمها في أبحاثها المختلفة.
- تعكف الوحدة على تطوير حزمة من الأدوات للمساعدة في عملية التحرير والنشر العلمي، وذلك في سياق مواكبة التحولات الكبيرة في صناعة النشر العلمي، عبر تأليل أجزاءٍ منها. وستُؤتمتُ هذه الأدوات مجموعةً من العمليات، تتمثّل في صياغة الهوامش المرجعية، ولوائح المراجع، وبناء فهارس الكتب وتصحيحها وتحريرها، إضافةً إلى الترجمة الآلية بأسلوب المركز العربي، وتوحيد ترجمة المصطلحات. وتأخذ الوحدة في صلب اهتمامها أن تكون الأدوات التي تنتجها ميسرة للاستخدام لدى الباحثين، لذلك ستعمل على استخدام واجهة تفاعلية مع تطبيق وورد (Word).
- طورت الوحدة آليات لجمع "البيانات الوصفية" (Metadata) الخاصة بالأوراق العربية العلمية الأكاديمية المنشورة، وطورت أيضًا آليات لاستخراج الاستشهادات من الأبحاث، وذلك لتسهيل استخدامها في الأبحاث البيبليوميترية حول المعرفة العربية في الحقول الأكاديمية المختلفة؛ ما سيمكن الباحثين من قياس المقالات العلمية باللغة العربية وفهمها وتبيُّن التأثير المتبادل فيما بينها.

خلاصة

تؤمن وحدة دراسة المجال الاجتماعي الرقمي العربي بأهمية التعاون مع مختلف المؤسسات البحثية والأكاديمية العربية والدولية؛ لتعزيز إنتاج المعرفة الرقمية حول المجتمعات العربية بأساليب علمية

9 Artur Kiulian et al., "From English-Centric to Effective Bilingual: LLMs with Custom Tokenizers for Underrepresented Languages," *arXiv*, 24/10/2024, accessed on 11/3/2025, at: https://acr.ps/1L9zQRC

10 Doha Albared, Hadi Hamoud & Fadi Zaraket, "Arabic Topic Classification in the Generative and AutoML Era," in: *The First Arabic Natural Language Processing Conference*, ArabicNLP 2023 (USA: Association for Computational Linguistics, 2023), pp. 399-404.

11 Osama Mraikhat, Hadi Hamoud & Fadi Zaraket, "AREEj: Arabic Relation Extraction with Evidence," in: *The Second Arabic Natural Language Processing Conference*, ArabicNLP 2024 (USA: Association for Computational Linguistics, 2024), pp. 67-72.

متقدمة. وتعاونت الوحدة مع "معهد قطر لأبحاث الحوسبة" (Qatar Computational Research Institute)، وشركة "غوغل"، تحت رعاية وزارة الاتصالات وتقنيات المعلومات في دولة قطر، للمساهمة في بناء "نموذج فنار اللغوي"[12] وتقييمه. وقد دُشّن هذا النموذج في نهاية عام 2024، خلال القمّة العالمية للذكاء الاصطناعي في قطر.

وتسعى الوحدة لتوفير أدوات وتقنيات حوسبية تساهم في تطوير العلوم الاجتماعية والإنسانية الرقمية، مستندةً إلى مقاربة تكاملية تجمع بين تحليل البيانات الضخمة والمعرفة الأكاديمية العميقة. وتعمل الوحدة أيضًا، بما تتيحه من خدمات بحثية وتدريبية، على تمكين الباحثين العرب من استثمار الفضاء الرقمي في دراسة الظواهر الاجتماعية والسياسية وتحليلها؛ ما يساهم في بناء محتوى عربي رصين ومؤثر في الأوساط الأكاديمية العالمية.

وتؤكد الوحدة التزامها برؤيتها في تطوير أدوات رقمية متخصصة تعكس خصوصية المجال العربي، وتساهم في تعزيز استقلال الباحثين والمؤسسات العربية معرفيًّا وتقنيًّا، وتدعو إلى توسيع آفاق التعاون بين مختلف الجهات الفاعلة في المجال الرقمي، سواء كانت أكاديمية أو بحثية أو تقنية، من أجل تحقيق نهضة علمية قائمة على المعرفة الحوسبية والتحليل الرقمي المتقدم. وبذلك، تتطلع الوحدة إلى أن تكون مركزًا مرجعيًا في مجال الدراسات الرقمية والاجتماعية العربية، وإلى أن تسهم في خلق بيئة بحثية ديناميكية تعزز قدرة المجتمعات العربية على مواكبة التطورات التكنولوجية وإنتاج المعرفة الرصينة.

لمتابعة وحدة دراسة المجال الاجتماعي الرقمي العربي، ينظر: https://www.acrps.ai

البريد الإلكتروني: dru@dohainstitute.edu.qa

12 ينظر: "فنار: مشروع الذكاء الاصطناعي العربي (النسخة التجريبية)"، شوهد في 2025/4/4، في: https://acr.ps/1L9zQjk

المراجع

العربية

المركز العربي للأبحاث ودراسة السياسات. **برامج ومشاريع**. توثيق الذاكرة الفلسطينية. في: https://acr.ps/1L9zR7R

المركز العربي للأبحاث ودراسة السياسات. **المؤشر العربي**. "برنامج قياس الرأي العام". في: https://acr.ps/1L9zRgX

المركز العربي للأبحاث ودراسة السياسات. **المؤشر العربي**. "برنامج قياس الرأي العام". في: https://acr.ps/1L9zQyC

المركز العربي للأبحاث ودراسة السياسات: في عشر سنين. الدوحة/ بيروت: المركز العربي للأبحاث ودراسة السياسات، تشرين الثاني/ نوفمبر 2022.

الأجنبية

The First Arabic Natural Language Processing Conference. ArabicNLP 2023. USA: Association for Computational Linguistics, 2023.

The Second Arabic Natural Language Processing Conference. ArabicNLP 2024. USA: Association for Computational Linguistics, 2024.