تطوير خدمات دعم التنقيب في النصوص والبيانات في مكتبة جامعية بحثية

ليزا بيرد، جيمس بارنيت

الملخص:

تعرض دراسة الحالة هذه كيف طورت خدمات المكتبة في جامعة برمنغهام دعمها للتنقيب في النصوص والبيانات بعد إدخال استثناء قانوني لذلك في المملكة المتحدة عام 2014. وتبدأ بمشروع في لسانيات المتون احتاج إلى أرشيف صحفي، فأسس شراكة بين فريق حقوق المؤلف والترخيص وفريق البنية التحتية البحثية في الجامعة. ثم تصف الفجوات التي كشفتها الاستفسارات، وهي: الوصول إلى البيانات، وإلى واجهات برمجة التطبيقات وتمويلها، وإلى المهارات. وتعرض مشروعًا شمل نقل مجموعات البيانات من الأقراص الصلبة إلى البنية التحتية، ومراجعة الأدوات، وإعداد صفحات إرشادية على الويب، واستبانة وُجهت إلى الباحثين أظهرت أن نقص المهارات وإتاحة البيانات أكبر العوائق. وتختتم بتأملات في مستقبل الخدمة: النسخ المقروءة آليًّا، وتطوير مهارات المكتبيين والباحثين، وتعزيز الصفحات الإرشادية، وأثر الذكاء الاصطناعي التوليدي، مؤكدة أن للمكتبات الأكاديمية دورًا في هذا المجال.

الكلمات المفتاحية: التنقيب في النصوص والبيانات، دعم البحث، المهارات البحثية، حقوق المؤلف والترخيص، تطوير الخدمات

مقدمة:

عُدّل قانون حقوق المؤلف والتصاميم وبراءات الاختراع لعام 1988 (Copyright, Designs and Patents Act 1988) عام 2014 ليتضمن المادة (29A): النسخ لأغراض تحليل النصوص والبيانات للبحث غير التجاري.[1] وتتيح هذه المادة الجديدة للباحثين إجراء تحليل حاسوبي للمصنفات التي يملكون حق الوصول المشروع إليها، متى كان التحليل لأغراض البحث غير التجاري حصرًا. ويناقش هذا المقال كيف طورت خدمات المكتبة في جامعة برمنغهام خدماتها لتلبية احتياجات الباحثين الذين أرادوا الاستفادة من هذا التغيير الجديد في القانون. ويتناول المقال المشاركة الأولى لخدمات المكتبة في مشروع للتنقيب في النصوص والبيانات، وكيف حُددت فجوات الخدمة في هذا المجال وعولجت، ويقدم تأملًا في الكيفية التي قد تتطور بها الخدمة في المستقبل.

الخلفية:

بعد عام من تغيير التشريع، تلقت خدمات المكتبة أول استفسار يتعلق بالتنقيب في النصوص والبيانات من أكاديمي كان ينضم إلى الجامعة ومعه مشروع ممول في لسانيات المتون يحلل استخدام اللغة في الأدب. وفي إطار المشروع، كان يحتاج إلى الوصول إلى أرشيف صحفي، وفهم المسائل القانونية ومسائل حقوق المؤلف، وتطوير حل للوصول إلى المواد واستضافتها. وقاد رئيس قسم حقوق المؤلف والترخيص، الذي يعمل ضمن خدمات المكتبة، عملية الدعم. وكانت المسائل التي ينبغي معالجتها:

- الوصول المشروع إلى المحتوى
- البنية التحتية لتقنية المعلومات لاستضافة المحتوى
- أمن البيانات وحفظها (preservation)
- كيفية مشاركة البيانات داخليًّا وخارجيًّا مع المتعاونين في البحث
- ما إذا كان آخرون سيحتاجون إلى الوصول إلى المواد في مشروعات بحثية مستقبلية.

اشتُري قرص صلب يحتوي على الأرشيف الصحفي. وكان الناشر على علم بخطة استخدام المحتوى، وروجعت الرخصة لمعرفة مدى توافقها مع الاستثناءات التشريعية الجديدة. وقد أصبح ذلك منذئذ جزءًا من مسار العمل المعتاد عند شراء مصادر جديدة، إذ يُعترض على الشروط المتعلقة بالتنقيب في النصوص والبيانات إذا حاولت تقييد حقوق الباحثين. ونحن نفضل أن تعكس الرخص التي نوقعها القانون بدلًا من الاضطرار إلى الاعتماد على حكم إبطال الشروط التعاقدية المخالفة (‘no contract’ override provision).[2] وعملت خدمات المكتبة مع فريق البنية التحتية البحثية في الجامعة،[3] وحُمّلت المواد على البنية التحتية البحثية لتمكين الأكاديمي من الوصول إليها. وكان من الممكن منح باحثين آخرين حق الوصول عند الطلب. ومنذ ذلك الحين عملت خدمات المكتبة مع فريق البنية التحتية البحثية على إنشاء استمارة إلكترونية تتيح للباحثين الآخرين في الجامعة طلب نسختهم الخاصة من الأرشيف على البنية التحتية البحثية. وقد حل وجود البيانات في هذه البيئة كثيرًا من المسائل الأولية المذكورة أعلاه. ويدير فريق حقوق المؤلف والترخيص في خدمات المكتبة أذونات الوصول.

وهذا يعني أن لدينا الآن نسختين على الأقل من ذلك الأرشيف على البنية التحتية:

- الأرشيف الأصلي كما حُمّل على البنية التحتية البحثية. وهي نسخة يمكن للباحثين الآخرين في الجامعة طلب نسخة منها لاستخدامها في بحوثهم الخاصة. وهذه هي «النسخة المرجعية الأصلية» (‘golden copy’).[4]
- النسخة التي عالجها فريق بحث لسانيات المتون بالتعرف الضوئي على الحروف (optical character recognition, OCR) وصححها في صورة ملف نصي (.txt). وكان إنجاز ذلك يدويًّا سيستغرق ساعات عمل كثيرة.

تحديد الفجوة في الخدمة:

بمجرد أن بدأ العمل المرتبط بمشروع لسانيات المتون، شرعنا في التفكير على نحو أوسع في الدور الذي قد تؤديه خدمات المكتبة في دعم أنشطة التنقيب في النصوص والبيانات لدى المجتمع البحثي الأوسع. ومن خلال الرد على الاستفسارات المتفرقة المتعلقة بالتنقيب التي وردت لاحقًا، تمكنا من تحديد فجوات ينبغي سدها (إما من قبل المكتبة وإما من قبل خدمات جامعية أخرى). وكانت الاستفسارات نفسها تصدر عادة عن باحثين يرغبون في استيضاح الآثار المترتبة، من حيث حقوق المؤلف والترخيص، على إمكانية التنقيب في محتوى بعض قواعد البيانات والأرشيفات التي نشترك فيها. وكانت طبيعة المحتوى متنوعة، وشملت أرشيفات صحفية وقضايا قانونية وبيانات

استشهادات. غير أنه اتضح من خلال التعامل مع هذه الاستفسارات أن ثمة مسائل مشتركة تبرز مرارًا، وتصبح بعد ذلك عوائق تحول دون تمكن الباحث (أو الباحثين) من إجراء أعمال التنقيب في النصوص والبيانات. وقد أتاحت لنا هذه المسائل رسم صورة للفجوات التي ينبغي سدها في الخدمة لإنشاء خدمة للتنقيب في النصوص والبيانات قابلة للتوسع. ونناقش هذه الفجوات فيما يلي.

الوصول إلى البيانات:

وردت استفسارات حول دعم التنقيب في النصوص والبيانات يتوجه فيها الباحث إلى خدمات المكتبة راغبًا في معرفة البيانات القابلة للتنقيب المتاحة له عبر اشتراكاتنا الحالية. فعلى سبيل المثال، توجه طالب دكتوراه إلى خدمات المكتبة طلبًا للدعم في مشروع يتطلب الوصول إلى أرشيفات صحفية، لكنه كان يحتاج أولًا إلى استيضاح الأرشيفات الصحفية (والمنصات التي تستضيفها) المتاحة في مجموعاتنا. وبعد استيضاح ذلك، دار حوار حول مدى إتاحة تلك المصادر لأنشطة التنقيب في النصوص والبيانات. فمثلًا: أي المصادر تتوافر لها واجهة برمجة تطبيقات (application programming interface, API) مجانية؟ وهل للمنصات حدود للتنزيل ينبغي مراعاتها؟ وهذا مجال واضح ينبغي للمكتبة أن تكون رائدة فيه، بالنظر إلى أن الترويج للمصادر وترسيخ استخدامها مبدأ أساسي في أي خدمة مكتبية، لكن الفجوة هنا تكمن في توضيح لا أين توجد البيانات فحسب، بل مدى إتاحة تلك البيانات لأنشطة التنقيب في النصوص والبيانات أيضًا.

الوصول إلى واجهات برمجة التطبيقات وتمويلها:

مع أن خدمات المكتبة، عبر فريق حقوق المؤلف والترخيص، في موقع جيد يتيح لها أن توضح للباحثين ما يمكن فعله وما لا يمكن فعله فيما يتعلق بالتنقيب في النصوص والبيانات في المصادر التي تشترك فيها، فإن أكبر عائق أمام الباحثين الذين يمارسون أنشطة التنقيب لا يكون في الغالب ما يمكن فعله، بل كيفية فعله. فعلى سبيل المثال، مع أن الباحث الذي يريد التنقيب في قاعدة بيانات أو أرشيف قد يتصل بالمكتبة في البداية للتأكد من أن رخصتنا تسمح له بإجراء أعمال التنقيب، فإنه كثيرًا جدًّا ما يُتبع هذا الاستفسار بأسئلة عما إذا كانت المكتبة قادرة على توفير الوصول إلى واجهات برمجة التطبيقات التي ينشئها الناشرون أو مزودو المحتوى لهذا العمل. ويطلب كثير من المزودين رسومًا إضافية مقابل الوصول إلى واجهات برمجة التطبيقات الخاصة بهم، وكثيرًا ما تكون باهظة إلى حد يحول دون الاستفادة منها (وتكلف أحيانًا أكثر من الاشتراك السنوي في قاعدة البيانات نفسها). وقد يستطيع الباحث الحاصل على منحة، تُدرج فيها تكلفة الوصول إلى واجهة برمجة التطبيقات المعنية ضمن تكاليف المشروع، تحمّل النفقات، لكن كيف نقدم خدمة منصفة لمن لا يحظون بتمويل؟ وبالنسبة إلى المكتبة، يصبح الاستثمار في واجهات برمجة تطبيقات باهظة الثمن خيارًا استراتيجيًّا صعبًا، إذ يجب موازنة الإنفاق المالي مع طلبات الحصول على مصادر أخرى.

الوصول إلى المهارات والبنية التحتية للدعم:

نشأت بعض الاستفسارات المتفرقة عن الحاجة إلى دعم لتطوير المهارات اللازمة لإجراء أنشطة التنقيب في النصوص والبيانات المطلوبة بنجاح. فعلى سبيل المثال، حيثما توافرت واجهة برمجة تطبيقات، قد يلزم بعد ذلك دعم لفهم كيفية الاستعلام منها للحصول على البيانات المطلوبة. وفضلًا عن ذلك، حيثما كانت مجموعات بيانات كبيرة متاحة لأغراض التنقيب دون واجهة برمجة تطبيقات، يحتاج بعض الباحثين إلى دعم في تطوير مهارات البرمجة لكتابة

واجهة برمجة التطبيقات الخاصة بهم. وهذه ليست مهارات نستطيع تقديم الدعم والتدريب فيها في خدمات المكتبة، ولذلك نعتمد على توجيه الباحثين إلى الدعم الذي يقدمه زملاؤنا في فريق البنية التحتية البحثية، الذين ينظمون تدريبًا على لغات برمجة مثل (R) و(Python).

مشروع التنقيب في النصوص والبيانات:

بعد تحديد الفجوات، بدأت خدمات المكتبة مشروعًا صغيرًا لدراسة ما يلي:

- مجموعات البيانات الأخرى التي لدينا على أقراص صلبة، وما إذا كان بإمكاننا إضافتها إلى البنية التحتية البحثية لحفظها واستخدامها في التنقيب مستقبلًا
- مراجعة أدوات التنقيب في النصوص والبيانات الجاهزة
- إعداد إرشادات على الويب بشأن التنقيب في النصوص والبيانات
- الدعم الذي يحتاج إليه الباحثون.

مجموعات البيانات على الأقراص الصلبة:

حددت خدمات المكتبة عدة أقراص صلبة تحتوي على مجموعات أرشيفية سبق شراؤها. وقد أُضيفت هذه إلى البنية التحتية البحثية. ونحن بصدد إعداد استمارات الطلب الإلكترونية والآلية اللازمة لتمكين الباحثين من الوصول إلى هذه «النسخ المرجعية الأصلية» لأغراض التنقيب، استنادًا إلى تجربتنا مع الأرشيف الصحفي.

أدوات التنقيب في النصوص والبيانات:

نظرت خدمات المكتبة أيضًا في اقتناء حق الوصول إلى أدوات للتنقيب في النصوص والبيانات، لكن لا يوجد حاليًّا مخصص مالي لتوفيرها. وقد تقدم أكاديمي في لسانيات المتون بطلب للحصول على أداة تمكّن طلابه من العمل بسهولة على مشروعات صغيرة للتنقيب. فاختارت خدمات المكتبة (Gale Digital Scholar Lab)،[5] الذي يتيح التنقيب في البيانات عبر كثير من الأرشيفات التي سبق لخدمات المكتبة شراؤها.

صفحات الويب الخاصة بالتنقيب في النصوص والبيانات:

أعد فريق حقوق المؤلف والترخيص المسودة الأصلية لصفحات ويب خاصة بالتنقيب في النصوص والبيانات لإرشاد الباحثين الراغبين في المشاركة في هذا النوع من البحث. وراجع فريق المهارات البحثية وفريق البنية التحتية البحثية الصفحات، وقدما ملاحظات حول كيفية دمج المعلومات عن خدماتهما فيها. وأتاح ذلك للصفحات أن تعكس، من مكان واحد، الدعم المركزي الأوسع المتاح للتنقيب في النصوص والبيانات على مستوى الجامعة. وتُستضاف الصفحات الآن على موقع خدمات المكتبة،[6] وتقدم معلومات عن:

- ماهية التنقيب في النصوص والبيانات ومسائل حقوق المؤلف المرتبطة بهذا العمل.
- المصادر المتاحة لدينا بالفعل على البنية التحتية لأغراض التنقيب. إضافة إلى تفاصيل عن وصولنا إلى (Gale Digital Scholar Lab) والأرشيفات المتاحة لنا للتنقيب ضمن ذلك المصدر.
- التنقيب في النصوص والبيانات والأدوات البرمجية، بما فيها البرمجيات المدفوعة والبرمجيات الحرة ومفتوحة المصدر (free and open source software, FOSS)

- إرشادات حول كيفية الحصول على دعم في البرمجة في الجامعة.

استبانة التنقيب في النصوص والبيانات:

أجرت خدمات المكتبة استبانة للعاملين وطلاب الدكتوراه المنخرطين في التنقيب في النصوص والبيانات، لفهم المزيد عن استخدام الباحثين الحالي للتنقيب والمشكلات التي واجهوها.

وحُدد الباحثون إما لأن خدمات المكتبة كانت قد تواصلت معهم من قبل بسبب استفسارات تتعلق بالتنقيب، وإما لأنهم مدرجون على صفحات الجامعة على الويب بوصفهم أعضاء في مجموعات بحثية مهتمة بالتنقيب في النصوص والبيانات، مثل مركز أبحاث المتون (Centre for Corpus Research)[7] ومعهد علوم البيانات والذكاء الاصطناعي متعدد التخصصات (The Institute for Interdisciplinary Data Science and AI).[8] وحُدد 150 شخصًا، تلقى كل منهم رسالة بريد إلكتروني شخصية تطلب منه استكمال الاستبانة. وكانت الاستبانة قصيرة، وكان غرضها معرفة:

- ما إذا كان هؤلاء الأشخاص يشاركون بالفعل في أنشطة التنقيب في النصوص والبيانات
- الأدوات التي يستخدمونها لإنجاز ذلك العمل
- ما إذا كانوا يواجهون أي عوائق في إجراء بحوث التنقيب (أو في التخطيط لإجرائها)
- مدى فائدة صفحات الويب الجديدة لخدمات المكتبة الخاصة بالتنقيب في نظرهم.

بلغ معدل الاستجابة للاستبانة 15%، وجاءت الردود من تخصصات مختلفة في الجامعة. وقال 78% من المستجيبين إنهم منخرطون في مشروعات للتنقيب في النصوص والبيانات. أما غير المنخرطين فذكروا أنهم يريدون إجراء بحوث تتضمن التنقيب لكنهم يحتاجون إلى تدريب.

واستخدم المستجيبون طائفة متنوعة من الأدوات والمهارات لإجراء بحوثهم في التنقيب، شملت مهارات البرمجة والبرمجيات والأجهزة. وحددوا 35 أداة مختلفة، لكن معظمها لم يكن يستخدمه إلا شخص أو شخصان.

واستخدم 39% من المستجيبين (R) و(RStudio)، وجاءت (Python) بعدهما بفارق محدود (ذكرها 30% من المستجيبين). وكان بعض المستجيبين يستفيدون من كل من (R) و(Python)، فيستخدمون (Python) للمعالجة المسبقة للبيانات و(R) لإتمام التحليل. وكان من المفيد معرفة لغة البرمجة التي يستخدمها الباحثون في هذه المشروعات، حتى تعرف خدمات المكتبة أين تركز جهودها إذا احتاج العاملون إلى الإلمام بهذه اللغات لتحسين دعمهم لباحثي التنقيب في النصوص والبيانات.

واتضح أن الأدوات الجاهزة لم تكن تلبي احتياجات الباحثين دائمًا، إذ أنشأ 13% من المستجيبين أدوات داخلية خاصة بهم. ويكشف ذلك أن الباحثين قد يحتاجون إلى قدر معين من المهارة التقنية والبرمجية عند إجراء مشروعات التنقيب.

وكان أكبر عائقين أمام إجراء بحوث التنقيب في النصوص والبيانات (وقد ذكر كلًّا منهما 17% من المستجيبين) هما المهارات/المعرفة التقنية وإتاحة البيانات. وعند النظر في التعليقات المتعلقة بالمهارات/المعرفة التقنية، اتضح أن الباحثين واجهوا صعوبات بسبب عدم امتلاكهم مهارات برمجة كافية، وأنهم لم يكونوا متأكدين كذلك من أفضل الأساليب لتنقية أي بيانات تمكنوا من جمعها. وبسبب العدد الكبير من الأدوات المتاحة على الإنترنت، واجهوا

صعوبة أيضًا في تحديد أكثرها فائدة لاحتياجاتهم، وأرادوا مزيدًا من الدعم التوجيهي. أما العوائق الرئيسية المتعلقة بالبيانات فكانت، على نحو غير مفاجئ، مرتبطة بالحاجة إلى تمويل لتغطية تكاليف الوصول إلى واجهات برمجة التطبيقات. وكان ذلك مصدر قلق خاص لمن لديهم بحوث غير ممولة. غير أن بعض المستجيبين لم يكونوا يعرفون ببساطة مجموعات البيانات المتاحة في الجامعة لبحوث التنقيب، وهو ما يبرز الحاجة إلى ترويج أفضل لصفحات الويب الخاصة بالتنقيب في النصوص والبيانات.

ورأى المستجيبون أن صفحات الويب الجديدة الخاصة بالتنقيب نقطة انطلاق جيدة لأي شخص جديد على هذا المجال، لكن يمكن تحسينها بتضمين دراسات حالة من باحثين يمارسون التنقيب حاليًّا، لإظهار إمكانات التنقيب في النصوص والبيانات بوصفه منهجًا بحثيًّا. كما طلبوا إضافة مزيد من الإرشادات المتعلقة بالمهارات إلى الصفحات.

تأملات وآفاق مستقبلية:

أبرز مشروع التنقيب في النصوص والبيانات - ولا سيما استبانة التنقيب - أنه رغم أن خدمات المكتبة تقدم الكثير لمستخدميها من حيث دعم التنقيب، فإن ثمة مجالًا لتوحيد ما تقدمه مع ما تقدمه الخدمات الأخرى في الجامعة وتطويره أكثر. ونتناول فيما يلي بعض هذه المجالات المستقبلية للتطوير.

نسخ البيانات المعالجة بالتعرف الضوئي على الحروف/المقروءة آليًّا:

أنشأ مشروع لسانيات المتون نسختين من الأرشيف الصحفي: «نسخة مرجعية أصلية» ونسخة بصيغة (TXT) عولجت بالكامل بالتعرف الضوئي على الحروف وصُححت. والملف النصي قد يكون أكثر فائدة للبحث لأنه جاهز للتنقيب. غير أن الباحثين قد يحتاجون إلى تحرير أي من النسختين لتلبية احتياجاتهم الخاصة. وإذا أُتيحت نسخ غير النسخة المرجعية الأصلية، فينبغي أن تصحبها معلومات عن كيفية تحريرها. والمثالي أن يقدم الناشرون نسخًا من أرشيفاتهم مصححة بعد التعرف الضوئي على الحروف، وهو ما سيوفر الوقت على مستوى القطاع بأكمله.

التعليم وتطوير المهارات في مجال التنقيب في النصوص والبيانات:

أظهرت استبانة التنقيب أن 22% من المستجيبين أرادوا المشاركة في بحوث التنقيب في النصوص والبيانات لكنهم سيحتاجون إلى تدريب، وهو ما يحدد فجوة واضحة. ومن الناحية الواقعية، سيكون رفع مهارات العاملين في المكتبة إلى مستوى مهارات البرمجة المتقدمة مسعى كبيرًا طويل المدى. غير أن تزويد العاملين الذين يُرجح أن يتعاملوا مع استفسارات التنقيب بمعرفة أساسيات البرمجة سيكون استثمارًا استراتيجيًّا مفيدًا. وسيكون هؤلاء العاملون عندئذ في موقع مثالي لفهم طبيعة استفسارات التنقيب فهمًا عميقًا وتحسين تجربة الباحثين، بتوجيه الاستفسارات بسلاسة بين العاملين في المكتبة ودعم البرمجة المتاح خارج المكتبة.

وثمة إمكانية لأن يستخدم فريق المهارات البحثية أداة مثل (Gale Digital Scholar Lab) - التي تتيح إجراء التنقيب في المصادر المتضمنة فيها دون حاجة الباحثين إلى امتلاك مهارات برمجة متقدمة - أساسًا لإعداد مواد تدريبية تعرّف بالتنقيب في النصوص والبيانات للباحثين غير الملمين بالمفهوم وغير المتأكدين من صلته ببحوثهم. والأمل هنا أن تؤدي خدمات المكتبة دورًا فاعلًا في توسيع الوعي بتقنيات التنقيب وتعزيز جاذبيتها لدى المجتمع البحثي الأوسع.

مواصلة تطوير صفحات الويب الخاصة بالتنقيب في النصوص والبيانات والترويج لها:

أظهرت استبانة التنقيب أن صفحات الويب الجديدة قدمت لكل من هو جديد على التنقيب في النصوص والبيانات معلومات مفيدة للبدء، لكن يمكن تحسينها بإضافة دراسات حالة وإرشادات بشأن المهارات وتوضيح مجموعات البيانات المتاحة. وإضافة دراسات الحالة تطوير مؤكد ينبغي لخدمات المكتبة العمل على تحقيقه. غير أن صفحات الويب الخاصة بالتنقيب تتضمن بالفعل معلومات عن فرص تطوير المهارات ومجموعات البيانات والأدوات المتاحة. ويشير ذلك إلى أن العمل المطلوب هو ضمان أن تكون المعلومات المنقولة عبر صفحات الويب واضحة وسهلة الوصول قدر الإمكان، ثم تعزيز انتشار هذه المعلومات في المجتمع البحثي عبر ترويج موجّه.

تطورات الذكاء الاصطناعي:

من الضروري أيضًا أن تكون خدمات المكتبة على وعي بالأثر الذي قد يحدثه في أنشطة التنقيب في النصوص والبيانات التسارع الكبير في تطورات الذكاء الاصطناعي، ولا سيما الذكاء الاصطناعي التوليدي (generative AI, GAI)، في أعقاب إطلاق (ChatGPT) في نوفمبر 2022.[9] فإذا أصبح استخدام أدوات الذكاء الاصطناعي التوليدي أكثر شيوعًا في العمليات البحثية،[10] فثمة إمكانية لاستخدام تقنية الذكاء الاصطناعي التوليدي في الاستعلام من مجموعات كبيرة من البيانات باستخدام التلقين باللغة الطبيعية (natural-language prompting) لإنجاز مهام تنقيب كانت تتطلب من قبل الوصول إلى واجهات برمجة التطبيقات أو مهارات برمجة متقدمة. وقد يفتح ذلك باب التنقيب في النصوص والبيانات أمام شريحة أوسع من الباحثين مما كان عليه الحال من قبل. غير أن هذه الإمكانية محفوفة أيضًا بالمخاطر - من مسائل قابلية إعادة إنتاج النتائج، وفهم مجموعة النصوص التي دُرّبت عليها الأداة («الصندوق الأسود»)، إلى احتمال التعدي على حقوق المؤلف[11] - ولذلك سيتعين على خدمات المكتبة أن تكون مستعدة لتثقيف المجتمع البحثي بشأن الفرص والمخاطر المصاحبة لاستخدام الذكاء الاصطناعي، سواء في أنشطة التنقيب في النصوص والبيانات أو بوصفه جزءًا من البحث على نحو أعم.

الخاتمة:

يمثل التنقيب في النصوص والبيانات، بوصفه منهجًا بحثيًّا، مجالًا يشهد نموًّا حقيقيًّا. فالتقدم في البيانات الضخمة وتعلم الآلة، إلى جانب صعود تخصصات مثل الإنسانيات الرقمية، يعني أن الباحثين يهتمون على نحو متزايد بكيفية تطبيق تقنيات التنقيب في مجالاتهم البحثية.[12] غير أنه تجدر الإشارة إلى أن دعم التنقيب في النصوص والبيانات لن يقتصر بالضرورة على المجتمع البحثي للمكتبة. ففي جامعة برمنغهام بدأنا نتلقى استفسارات بشأن التنقيب من طلاب المرحلة الجامعية الأولى، ومن أكاديميين يدمجون التنقيب في ممارساتهم التدريسية. ولذلك فإن الدعم الذي تطوره المكتبة في مجال التنقيب في النصوص والبيانات سيحتاج على الأرجح إلى أن يشمل دعم التعليم والتعلم أيضًا.

وتُظهر تجربتنا في دعم التنقيب في النصوص والبيانات أن للمكتبات الأكاديمية دورًا في هذا المجال. فبعض الدعم المطلوب للتنقيب امتداد طبيعي للدعم الذي تقدمه المكتبات بالفعل لمجتمعاتها. ويشمل ذلك المشورة بشأن حقوق المؤلف والترخيص فيما يتعلق بالتنقيب، إلى جانب الترويج للمصادر ومجموعات البيانات المتاحة لأغراض التنقيب، ومهارات الثقافة المعلوماتية اللازمة لتقييم مدى ملاءمتها.

غير أن دعم أنشطة التنقيب في النصوص والبيانات سيتيح للمكتبات أيضًا فرصًا لتطوير مهاراتها ودعمها. ويشمل ذلك أن يطور المكتبيون مهارات تمهيدية في البرمجة واستخدام واجهات برمجة التطبيقات لضمان أقصى قدر من الكفاءة في التعامل مع استفسارات التنقيب. ويشمل كذلك أن يكونوا في طليعة من يثقفون الباحثين بشأن

الثقافة اللازمة لمواجهة الفرص والمخاطر التي ستحملها تطورات الذكاء الاصطناعي في سياق التنقيب في النصوص والبيانات (وفي سياق البحث الأوسع). ونحن نرحب بفرص التدريب التي تقدمها (CILIP) و(RLUK) و(Jisc)، [13] المتاحة بالفعل للمكتبيين الذين يتطلعون إلى البدء في سد تلك الفجوة، لكن سيتعين بذل المزيد لضمان امتلاك المكتبيين العاملين في المكتبات البحثية المهاراتِ التي يحتاجون إليها لدعم العاملين والطلاب في عملهم في التنقيب في النصوص والبيانات.

المراجع:

1. "Copyright, Designs and Patents Act, 1988, c. 48, s. 29A," legislation.gov.uk, https://www.legislation.gov.uk/ukpga/1988/48/section/29A (accessed 23 January 2024).
2. legislation.gov.uk, "Copyright, Designs and Patents Act, 1988, c. 48, s. 29A (5)."
3. "Birmingham Environment for Academic Research (BEAR)," University of Birmingham, https://www.birmingham.ac.uk/research/arc/bear/index.aspx (accessed 23 January 2024).
4. "What is a golden copy?," University of Edinburgh, https://www.ed.ac.uk/records-management/guidance/records/retention/golden-copy (accessed 23 January 2024).
5. "Gale Digital Scholar Lab: Open New Research Pathways," Gale, https://www.gale.com/intl/primary-sources/digital-scholar-lab (accessed 23 January 2024).
6. "Text and data mining," University of Birmingham Intranet, https://intranet.birmingham.ac.uk/as/libraryservices/library/copyright/text-and-data-mining/text-and-data-mining.aspx (accessed 23 January 2024).
7. "Centre for Corpus Research," University of Birmingham, https://www.birmingham.ac.uk/research/activity/corpus/index.aspx (accessed 23 January 2024).
8. "The Institute for Interdisciplinary Data Science and AI," University of Birmingham, https://www.birmingham.ac.uk/research/data-science/index.aspx (accessed 23 January 2024).
9. Bernard Marr, "A short history of ChatGPT: how we got to where we are today," Forbes, May 19, 2023, https://www.forbes.com/sites/bernardmarr/2023/05/19/a-short-history-of-chatgpt-how-we-got-to-where-we-are-today/ (accessed 23 January 2024).
10. Jack Grove, "The ChatGPT revolution of academic research has begun," Times Higher Education, March 16, 2023, https://www.timeshighereducation.com/depth/chatgpt-revolution-academic-research-has-begun.
11. Alex Fenlon, "AI: what are the risks?," UoB PGR Development (blog), June 26, 2023, https://uobpgrdevelopment.wordpress.com/2023/06/26/ai-what-are-the-risks/ (accessed 23 January 2024).

12. Peter Findlay, "Why should libraries publish AI-ready collections?," Jisc Blog, November 9, 2023, https://beta.jisc.ac.uk/blog/why-should-libraries-publish-ai-ready-collections (accessed 23 January 2024); Andrew Cox, Research report: the impact of AI, machine learning, automation and robotics on the information profession (London: CILIP, 2021), https://www.cilip.org.uk/page/researchreport (accessed 23 January 2024).

13. "UKeiG CPD Workshop: Artificial intelligence for librarians, information and knowledge professionals," CILIP, https://www.cilip.org.uk/events/EventDetails.aspx?id=1772016&group= (accessed 23 January 2024); "RLUK Digital Shift Forum," RLUK, https://www.rluk.ac.uk/dsf/ (accessed 16 November 2023); "Artificial intelligence and ethics," Jisc, https://beta.jisc.ac.uk/training/artificial-intelligence-and-ethics (accessed 23 January 2024).