

خطة إدارة البيانات

Data Management Plan

اسم المشروع: "AI- Muqtafi Smart Search"

المشكلة: تتمثل الإشكالية بآلية البحث المعتمدة في المقتفي، فهي تعتمد على حرفية الكلمات المكتوبة بالبحث وليس على المعنى المقصود منها، وبالتالي صعوبة الحصول على النتيجة المطلوبة في حال عدم معرفة المصطلحات القانونية أو اسم التشريع المطلوب.

الحل: تطوير عملية البحث في المقتفي بالاعتماد على "Elasticsearch" و "Natural Language Processing" ليتمكن المستخدم من البحث بلغة بسيطة والحصول على نتيجة دقيقة، هذا كخطوة أولى في سبيل تطوير عملية البحث على المقتفي. والخطوة الثانية تتمثل بتطوير عملية البحث بما يمكن المستخدم من طرح أسئلة قانونية والحصول على المواد القانونية والأحكام القضائية التي قد يستدل منها على جواب السؤال المطروح وهذا باستخدام "Question answering".

1. Data & Data Collection:

يوجد في المقتفي بيانات متعددة والتي يتم تحديثها باستمرار وفقاً للتشريعات والأحكام القضائية التي تصدر في فلسطين، تتمثل هذه البيانات بما يلي:

1. قاعدة التشريعات "Databases": وهي التي تحتوي على جميع التشريعات التي سنت في فلسطين من فترة الحكم العثماني حتى آخر التشريعات الصادرة عن السلطة الوطنية الفلسطينية. جُمعت هذه التشريعات من مصادرها الرسمية، والتي تتمثل حالياً بجريدة الوقائع الفلسطينية الصادرة عن ديوان الفتوى والتشريع، والتي تصدر بشكل دوري كل شهر.

تخزن هذه التشريعات في المقتفي على صيغة PDF و XML وتظهر للمستخدم بصيغة (HTML & PDF). حيث يكون التشريع الذي على صيغة PDF عبارة عن صورة التشريع كما نشر في جريدة الوقائع الفلسطينية، دون أي إضافات. أما التشريع على صيغة XML فيتم إضافة الفهرس للتشريع من خلال عنوان المواد التي يحتويها، وذلك لتسهيل قراءته والبحث فيه. تتم عملية الفهرسة "Data Annotation" يدوياً من خلال الدائرة القانونية في معهد الحقوق بجامعة بيرزيت.

يمكن الوصول إلى التشريعات في هذه القاعدة من خلال تسجيل الدخول في المقتفي والبحث في التشريعات المنفردة، أو البحث في التشريعات المدمجة أو الأحوال الشخصية للمسيحيين. ويختلف البحث في التشريعات المدمجة عن

البحث في التشريعات المنفردة، بأن نتيجة البحث في التشريعات المدمجة تكون عبارة عن التشريع مع جميع التعديلات التي طرأت عليه، في ملف XML ويظهر للمستخدم HTML، حيث تتم عملية دمج هذه التشريعات يدوياً من خلال الدائرة القانونية في معهد الحقوق.

تظهر نتيجة البحث في كلا الحالتين السابقتين، على هيئة جدول يحتوي على عنوان التشريع والحقبة التاريخية التي صدر بها وحالة التشريع، جميع هذه البيانات هي عبارة عن text ومخزنة XML. فإذا وجد المستخدم التشريع الذي يبحث عنه سينتقل إلى البطاقة التعريفية للتشريع والتي تحتوي على بعض البيانات الوصفية له مثل (حالة التشريع وتصنيفه وتاريخ نشره)، لينتقل فيما بعد إلى التشريع سواء PDF أو HTML.

2. قاعدة الأحكام النظامية: تحتوي على جميع الأحكام القضائية الصادرة عن المحاكم العليا الفلسطينية من العام 1994 حتى الآن. يتم الحصول على هذه الأحكام بالتعاون مع مجلس القضاء الأعلى، من خلال منح المعهد إمكانية الوصول إلى الميزان، بصورة دورية شهرية، تكون هذه الأحكام على هيئة Word. تخزن في المقتفي PDF وXML وتظهر للمستخدم (HTML & PDF). حيث يكون الحكم الذي على صيغة PDF عبارة عن صورة الحكم كما تم الحصول عليه من الميزان، دون أي إضافات. أما الحكم على صيغة XML فيتم فهرسة محتوياته لتسهيل قرائته والانتقال فيه لما يهم القارئ بسهولة، وتتم عملية الفهرسة "annotation data" يدوياً من خلال الدائرة القانونية في معهد الحقوق بجامعة بيرزيت.

يمكن الوصول إلى هذه الأحكام من خلال تسجيل الدخول والبحث في الأحكام أو البحث في الأحكام - كشاف مترابط أو استعراض الأحكام بالكشاف. تظهر نتيجة البحث في جميع الحالات السابقة على هيئة جدول يحتوي على الحكم والمحكمة والمدينة ونوع التقاضي والمجال. جميع هذه البيانات هي عبارة عن text ومخزنة XML.

فإذا وجد المستخدم الحكم الذي يبحث عنه سينتقل إلى بطاقة الحكم والتي تحتوي على بعض البيانات الوصفية له مثل (المحكمة ونوع التقاضي وتاريخ الفصل)، وتحتوي أيضاً على الحكم بصيغة PDF وHTML ومسارات الكشاف التي يمكن الوصول للحكم من خلالها. حيث يتم تصنيف كل حكم وفقاً لمحتوياته وللحلل القانوني الذي صدر فيه تحت ما يسمى بكشاف الأحكام، وهو عبارة عن مصطلحات دالة على الحكم "Index" يتم إدخالها يدوياً من الدائرة القانونية في معهد الحقوق، وتقييم أهمية كل حكم بالنظر إلى مشتملاته وأهميته من الناحية القانونية وهل إذا ما كان يستحدث مبدأ قانونياً جديداً أم لا؟ وتقوم الدائرة أيضاً بربط أحكام محكمة النقض بأحكام محاكم الدرجة الثانية، وربط كل حكم بالمبادئ القانونية التي يحتويها. تكون هذه البيانات على صيغة text ومخزنة XML.

3. قاعدة الأحكام الدينية: تحتوي على بعض الأحكام القضائية الصادرة عن المحاكم الشرعية العليا من العام 1989 حتى الآن. يتم الحصول على هذه الأحكام بالتعاون مع المحاكم الشرعية في المحافظات. تخزن في المقتفي PDF وXML وتظهر للمستخدم (HTML & PDF). حيث يكون الحكم الذي على صيغة PDF عبارة عن صورة الحكم دون أي إضافات. أما الحكم على صيغة XML فيتم فهرسة محتوياته لتسهيل قرائته والانتقال فيه لما يهم القارئ

بسهولة، وتتم عملية الفهرسة "annotation data" يدوياً من خلال الدائرة القانونية في معهد الحقوق بجامعة بيرزيت.

عملية البحث في الأحكام الدينية تم من خلال البحث في قاعدة المحاكم الدينية أو استعراض الأحكام الدينية بالكشاف، بطريقة مشابهة لطريقة البحث في الأحكام النظامية كما تم إيضاحه سابقاً.

4. ومن البيانات التي يحتوي عليها المقتفي أيضاً، والتي ستساعد في تطوير آلية البحث؛ المكنز القانوني: وهو عبارة عن المصطلحات القانونية التي تم تجميعها من النصوص الكاملة لبنك المعلومات القانونية لمعهد الحقوق. وتشتمل هذه المصطلحات على المصطلح القانوني الأعم والأضيق والمصطلحات المترابطة معه والمصطلح باللغة الإنجليزية، تكون هذه المصطلحات على صيغة text.

5. ومن البيانات التي يخزنها المقتفي، بيانات المستخدم "Logs". والتي تشمل بريد المستخدم الإلكتروني واسمه ورقمه والمؤسسة التي يعمل بها والغرض من الاستخدام. فمن خلال هذه البيانات يقوم المقتفي بتخزين تحركات المستخدم ومجالات بحثه والمدة التي يقضيها في كل بحث. تساعدنا هذه البيانات على تطوير نتيجة البحث لتصبح أكثر تعلقاً بموضوعه، إلا أن طريقة تسجيل هذه البيانات بحاجة إلى إعادة تنظيم وتخطيط بطريقة تمكننا من الاستفادة منها مستقبلاً.

من الممكن فيما بعد العمل على تطوير بيانات المستخدم، لتشمل فئة المستخدم والذي من الممكن أن يكون: طالب قانون أو باحث أو محامي أو قاضي... فهذه البيانات ستتمكننا من دراسة سلوك المستخدم عن طريق ربطها بتخصصه وتحركاته والاستفادة منها في تطوير البحث ونتيجته.

2. Standards and metadata:

تخزن Metadata للأحكام والتشريعات على أساس Structured Data عن طريق XML، وتختلف Metadata للتشريعات عنها للأحكام على النحو الآتي:

Metadata للتشريعات: وهي البيانات الوصفية المأخوذة من التشريع نفسه عند نشره، وتظهر هذه البيانات في البطاقة التعريفية للتشريع عند البحث عنه، وتحتوي على ما يلي:

- المصدر الذي أخذ منه التشريع، وهو بالعادة الجريدة الرسمية التي نشر بها في الحقبات المختلفة.
- تاريخ نشر التشريع بالميلادي والهجري، كما ورد في الجريدة الرسمية.
- عدد الجريدة الرسمية التي نشر فيها التشريع.
- نوع عدد جريدة الوقائع الفلسطينية التي نشر بها التشريع (عدد عادي أم ممتاز؟).

- مكان صدور التشريع.
- عنوان التشريع.
- سنة صدور التشريع.
- نوع التشريع.
- التصنيف الرئيسي للتشريع، وهذه البيانات تدخل من خلال الدائرة القانونية في المعهد.
- التصنيف المتفرع للتشريع.
- حالة التشريع؛ هل هو ساري أم ملغى؟
- مكان سريان التشريع.
- الحقبة التي صدر بها التشريع.
- تاريخ نفاذ التشريع، كما ورد فيه.
- رقم الصفحة التي نشر بها التشريع في الجريدة الرسمية.
- عدد صفحات التشريع كما نشر في الجريدة الرسمية.
- عدد المواد التي يحتوي عليها التشريع.
- ملاحظات على التشريع في حال وجدت يتم إدخالها من الدائرة القانونية بالمعهد.
- الجهة مصدرة التشريع.
- تاريخ صدور التشريع، كما ورد في خاتمة التشريع.

يتم إدخال هذه البيانات الوصفية من خلال الدائرة القانونية في معهد الحقوق، والتي تأخذ من التشريع نفسه ومن المصدر الذي نشر فيه التشريع، يتم إدخال هذه البيانات بشكل متزامن مع عملية إدخال التشريعات إلى المنظومة، أي بشكل شهري.

Metadata للأحكام: وهي البيانات الوصفية للأحكام القضائية والتي تأخذ منها، وتظهر هذه البيانات في بطاقة الحكم عند البحث عنه، وتحتوي على ما يلي:

- اللغة التي صدر بها الحكم.
- عنوان الحكم.
- المحكمة التي صدر عنها الحكم.
- الحقل الذي صدر فيه الحكم (مجال الحكم).
- نوع القضية (استئناف أو نقض...).
- مكان صدور الحكم.
- وجود رأس مخالف أم لا؟

- تاريخ صدور الحكم.
- المبادئ التي يحتوي عليها الحكم.
- التعليقات القانونية على الحكم القضائي.
- أهمية الحكم.
- رقم الحكم.
- السنة التي صدر بها الحكم.
- كشف الحكم، والذي يتم وضعه من الدائرة القانونية، بعد قراءة الحكم وتصنيفه.

يتم إدخال هذه البيانات الوصفية من خلال الدائرة القانونية في معهد الحقوق والتي يؤخذ بعضها من نص الحكم القضائي، والبعض الآخر يتم وضعه من الدائرة القانونية بعد قراءة الحكم وتحليله. تدخل هذه البيانات الوصفية بالتزامن مع عملية إدخال الأحكام بشكل دوري.

يعتمد المعهد في جمع جميع البيانات السابقة وإدخالها على Schema تسمى Akoma Ntoso وهو معيار تقني دولي مستخدم لتمثيل الوثائق التنفيذية والتشريعية والقضائية بطريقة منظمة باستخدام Legal XML Vocabulary.

3. Data Sharing:

إن البيانات التي يحتوي عليها المقتفي متاحة للجميع بغض النظر عن تخصصاتهم المهنية أو مجالات عملهم، ويمكن للجميع استخدامه في أي وقت وفي جميع الأماكن دون أي قيد، سوى تسجيل دخولهم إلى المقتفي، ولكن منعاً لتنزيل البيانات التي يحتويها المقتفي ونشرها في مواقع أخرى فقد وضع المقتفي حد أقصى للتشريعات والأحكام التي يمكن الإطلاع عليها وهي تقارب 80 حكم أو تشريع في اليوم الواحد.

ومن الأمور التي يقوم المعهد بتحديثها في المقتفي الجديد الذي لم يتم إطلاقه بعد، هو إتاحة إمكانية الوصول لبعض التشريعات المنفردة دون الحاجة إلى تسجيل الدخول إلى المقتفي، وذلك لزيادة أعداد مستخدميه. ففي الآونة الأخيرة، أصبح هناك العديد من المواقع التي تتيح هذه الخاصية للمستخدم. لذلك يعمل المعهد على تسهيل الوصول للتشريعات المنفردة دون الحاجة إلى تسجيل الدخول، بينما يُلزم المستخدم بتسجيل الدخول في حالة أراد البحث في التشريعات المدمجة والأحكام القضائية التي استحدثها المعهد ويتميز بها المقتفي عن غيره من المواقع.

من الممكن التفكير بخطوة مستقبلية، قد تساهم في تطوير المقتني ومساعدته على تحقيق عائد مادي، وهي إتاحة استخدام المقتني مجاناً في فلسطين، ولكن فرض رسوم اشتراكية سنوية لمستخدميه من الدول الأخرى وخاصة للمؤسسات التعليمية، ولتطبيق هذه الخطوة لابد ابتداء من أخذ موافقة جامعة بيرزيت.

Confidentiality and Ethics Issues:

من الأمور التي يسعى المعهد على المحافظة عليها هي خصوصية المستخدم والمعلومات التي يقوم بتسجيلها عند تسجيل الدخول كمستخدم جديد وهذه البيانات هي عبارة عن (اسم المستخدم وبريده الالكتروني والمؤسسة التي يعمل بها والدولة ورقم الهاتف ورقم الفاكس والغرض من الاستخدام). فقد ذكر في بند إخلاء المسؤولية بالمقتني أن "المعهد يبذل قصارى جهده للحفاظ على خصوصية المستخدمين ضمن الإمكانيات المتاحة وفي نطاق القانون". أما فيما يتعلق بأمن الشبكة والبريد الالكتروني واسم المستخدم وكلمة المرور فهم من مسؤولية المستخدم، وذلك كما ورد في اتفاقية الاستخدام.

ومن الأمور التي يجب العمل على تغييرها فيما يتعلق ببيانات المستخدم هو إتاحة الإمكانية للمستخدم لتغيير بياناته المدخلة أو إغلاق حسابه فيما بعد.

في النسخة الجديدة من المقتني فإن البيانات المطلوب إدخالها لتسجيل مستخدم جديد هي فقط، اسم المستخدم وبريده الالكتروني. ولكن اعتقد أنه من المهم الحصول على مجال عمل المستخدم وتخصصه، والتي يمكن الاستفادة منها فيما بعد لتطوير النتائج في عملية البحث وربطها بتخصص المستخدم وتحركاته السابقة.

Intellectual Property, Embargoes, and Protections:

يتمتع المقتني بالحماية القانونية لحق المؤلف وفقاً للقوانين المحلية والاتفاقيات الدولية، فكما ذكر في اتفاقية الاستخدام بأن معهد الحقوق في جامعة بيرزيت "يمتلك كل حقوق المؤلف المتعلقة بمنتج البرنامج، بما في ذلك دون حصر كل ما يدرج في منتج البرنامج من صور، أو نصوص مكتوبة، أو أدوات استخدام التطبيقات، والمطبوعات المرفقة به، وأية نسخ من البرنامج. أما نصوص التشريعات فلا تعتبر بحد ذاتها ملكاً للمعهد. ويعد البرنامج محمياً بالقوانين المحلية والإقليمية والاتفاقيات الدولية الخاصة بحق المؤلف. لذلك يعامل البرنامج كغيره من المصنفات المحمية بحقوق المؤلف".

وبناء على اتفاقية الاستخدام فإن المعهد يعطي المستخدم رخصة لاستخدام المقتني فقط، فقد ذكر بها ما يلي: " يتمتع معهد الحقوق، منتج البرنامج، بحماية القوانين المحلية والإقليمية والاتفاقيات الدولية الخاصة بحق المؤلف، علاوة على غيرها من القوانين والاتفاقيات الدولية في مجال الملكية الفكرية ولا يعد هذا المنتج مباعاً بل مرخصاً باستخدامه وفق هذه الوثيقة" وبالتالي فإن المقتني يتمتع بالحماية القانونية للملكية الفكرية وحقوق المؤلف.

إن أي مخالفة لاتفاقية الاستخدام أو أي من القوانين المحلية والاتفاقيات الدولية المتعلقة بحقوق المؤلف والملكية الفكرية، ستعرض صاحبها للمسائلة القانونية. بالإضافة إلى حق المعهد باتخاذ التدابير أو الإجراءات الدائمة أو المؤقتة التي يراها مناسبة من أجل حماية مصالحه، ومن هذه التدابير إنهاء رخصة المستخدم.

4. The guidelines, procedures, or policies for data reuse and/or redistribution, attribution, as well as for creation of derivatives from the data:

يعمل المعهد على تطوير وتحديث بيانات المقتني بشكل مستمر، وإيجاد روابط جديدة بين البيانات التي يحتويها بما يساعد المستخدم على الاستفادة من البيانات الموجودة بأكبر قدر ممكن، إلا أن فكرة مشاركة هذه البيانات مع جهة أخرى أو حتى منحها لجهة أخرى للعمل عليها، بحاجة إلى موافقة جامعة بيرزيت ومعهد الحقوق.

5. Long-Term Preservation and Archiving of Data:

تُخزن البيانات على سيرفرات المعهد، ويتم عمل نسخ احتياطية يومية لها.