الضبط المصطلحى داخل شبكة دولية متعددة المجالات*

"Vocabulary Control Within An International Multi - Disciplinary Network"

تأليف ولفرد لانكستر

ترجمة د. حسين الهبائلى

خبير التوثيق والمعلومات وأستاذ اللغويات المشارك

جامعة تونس الآلى

مستخلص

يعد الضبط المصطلحى مسألة هامة داخل شبكة دولية متعددة المجالات، لاسيما عند استخدام أكثر من لغة واحدة. والحلول الممكنة لهذه المسألة تتطلب: اعتبار كل المدخلات «لغة طبيعية» وتوفير وسائل بحث متنوعة، وبناء شبكة موحدة من المصطلحات، واستخدام المعاجم الوسيطة، وتبنى طريقة المكانز الكبيرة/ الصغيرة.

على أن كلاً من هذه المتطلبات تتضمن محاسن ومساوئ. والمسائل المصطلحية التى تواجهنا ليست يسيرة إلا أن التغلب عليها ليس مستحيلاً.

١ ـ مقدمة:

إن الدول العربية غنية بمصادر المعلومات، ومن الأدلة الواضحة على ذلك توافر مراكز المعلومات المتعددة المواضيع والمجالات. وللشبكة دور فى تنسيق أنشطة هذه المراكز وتحسين نوعية النفاذ إلى المعلومات فى كافة أنحاء الوطن العربى. ولكن هناك عدة مشاكل يجب مواجهتها عند تصميم الشبكة وتشغيلها، كما يجب تأمين الهيكل الأساسى الذى تبنى عليه ودمج أعمال المراكز القائمة. وإحدى هذه المشاكل هى مسألة الضبط المصطلحى. وتعالج هذه الورقة بعضاً من مشكلات الضبط المصطلحى داخل شبكة من نوع «أريسنات» (Arisnet) كما تقترح بعض الطرق الممكنة لحل هذه المشاكل. وتقتصر هذه الورقة على عرض افتراضين اثنين: أولهما، يربط مساهمة مراكز المعلومات داخل الشبكة بالحفاظ على قواعد البيانات فى شكل مقروء آلياً. وثانيهما يتطلب من هذه القواعد أن تقدم عدة طرق مختلفة لمعالجة الضبط المصطلحى، ولتوفير عدد من المكانز المختلفة، وخطط التصنيف، وقوائم رؤوس الموضوعات لتستخدم إلى جانب قواعد البيانات الخالية من الضبط المصطلحى (أى المبنية على اللغة الطبيعية).

والوضع الأمثل بالنسبة إلى الشبكة هو أن يتمكن كل مستفيد من النفاذ إلى السجلات

* ورقة عمل أعدت للمؤتمر الإقليمى الأول للشبكة العربية للمعلومات تونس – يونيو ١٩٨٧.

الببليوغرافية أو غيرها لأى مركز معلومات داخل الشبكة. ولتحقيق هذا الهدف تتوفر إمكانيتان اثنتان:

١ ـ دمج سجلات كافة المراكز داخل الشبكة الواحدة لقاعدة المعلومات.

٢ ـ المحافظة على هوية كل قاعدة من قواعد البيانات، وربط كل هذه المصادر مع بعضها البعض داخل الشبكة بواسطة الخط المباشر، وتمكين المستفيد فى أى مركز كان من النفاذ مادياً إلى محتوى قواعد البيانات المتوافرة بكل المراكز الأخرى.

٢ ـ توحيد قواعد البيانات:

(Integration of Data Bases)

لنفرض أنه تقرر دمج كل السجلات المتوافرة بكافة المراكز فى قاعدة واحدة للبيانات من نوع «أريسنات» فتكون المتطلبات الأولى فى هذه الحالة تطوير برامج تحول الأشكال العديدة للسجلات إلى شكل واحد يصلح للمعالجة الجماعية. وهذا يعنى أن كل السجلات يمكن التحكم فيها بواسطة مجموعة من البرامج. وعندما تكون قاعدة البيانات متوافرة فى شبكة من الحواسيب، أو مزدوجة الاستنساخ، أو مخزنة فى عدة أوعية، يمكن عندئذ لكل مستفيد من الشبكة النفاذ إلى أى مصدر من مصادر المعلومات المخزن بها باستخدام لغة بحث واحدة.

ورغم هذه الإمكانات فإن المستفيد لايزال يواجه مشكلة عظمى: هى أن كل قاعدة بيانات قد وقع تكشيفها بطريقة مختلفة. كما تختلف مجموعة المصطلحات من شبكة إلى أخرى. كيف يمكن حل هذه المشكلة؟ هناك عدة حلول ممكنة.

٢ . ١ ـ معالجة اللغة الطبيعية

(Natural Language Processing)

يتمثل الحل الأول فى التخلى عن استخدام اللغات المقيدة الموجودة، واعتبار قاعدة البيانات الموحدة ملفاً كبيراً من اللغة الطبيعية. وهذا يعنى أن كل عناصر البحث فى كافة مكونات قاعدة البيانات: (مثل الواصفات، ورؤوس الموضوعات، والكلمات المفتاحية فى العناوين والمستخلصات، وأسماء المؤلفين وغير ذلك) يجب دمجها فى ملف مقلوب. وبهذا يتمكن المستفيد من النفاذ إلى الملف الموحد وكأنه يبحث فى قاعدة عريضة للبيانات مثل قاعدة المستخلصات الكيمائية، وذلك بواسطة اللغة الطبيعية.

ما هى المشاكل التى تتجر للمستفيد عند البحث عن هذه المصادر؟

المشكل الأول هو التأكد من عدم ضياع المعلومات فى ملف يجمع بين مصطلحات مختلفة الأنواع. وعلى سبيل المثال إذا أدخل المستفيد مصطلحاً مثل: «Acid Rain» فإنه يجب أن يسترجع وحدات تندرج فيها هذه العبارة فى العناوين أو فى المستخلصات، بشرط أن لا تكون الوحدات التى ترد فيها هذه العبارة مجرد رأس موضوع، أو واصفة. وهذه المشكلة سهلة نسبياً ويمكن الاهتمام بها فى برامج البحث.

أما المشكلة الثانية فتتعلق بالالتباس الناتج عن التجانس الخطى. وتعد هذه المشكلة من أكبر المشاكل بالنسبة إلى قاعدة بيانات متكاملة ومتعددة المجالات، مثل التى تتطلبها «أريسنات»، خلافاً لما عليه الحال بالنسبة إلى الشبكة المخصصة لمجال واحد. غير أن هذه المشكلة ليست على هذه الصعوبة التى ظهرت عليها من أول وهلة فكلمة «PLANT» مثلاً التي تحمل في حد ذاتها لبساً، تفقد كثيراً من الالتباس بفضل السياق الذى ترد فيه مع مصطلحات أخرى عند استراتيجية البحث.

وهناك مشكلة أكثر أهمية تتمثل فى أن السجلات تستوجب على الأقل استخدام الفرنسية والإنجليزية إلى جانب العربية. وبينما يمكن دمج الفرنسية والإنجليزية فى هجائية واحدة فإن دمج اللغات ذات الكتابات المختلفة يعد مشكلة من نوع خاص تتعدى معالجتها نطاق هذه الورقة.

وتشكل قاعدة البيانات المتعددة المجالات مشكلتين بديهيتين: أولها مشكلة التجانس الخطى المتمثل فى الكلمات المركبة: فهناك مجموعة خاصة من الحروف تفيد معنى خاصاً بالنسبة للإنجليزية بينما تفيد معناً مخالفاً تماماً بالنسبة للفرنسية . وهذا ليس بالمشكل الذى لا يمكن تجاوزه، لا سيما وأن المصطلحات التى هى فى شكل ملف مقلوب يمكن تمييزها وفق المصدر المشتقة منه. وهكذا فإن الكلمة الفرنسية «Four» (فرن) يمكن الحفاظ عليها بعيدة عن الكلمة الإنجليزية «Four» (رقم). غير أن المحافظة على هذا التمييز تتطلب من النظام معرفة ما إذا كان مستخدم الشبكة يبحث بالإنجليزية أم بالفرنسية.

أما المشكلة الثانية والأكثر أهمية والمتمثلة فى أن قواعد البيانات المتعددة المجالات كثيراً ما تزيد فى حدود نسيان بعض الوحدات الملائمة.

وللمحافظة على نفس المثال، يستخدم الباحث مصطلح "Fours" إلى جانب "Furnaces" فى قاعدة بيانات تمزج بين مدخلات فرنسية وإنجليزية. وهذه المشكلة يمكن تفاديها بتكوين ثلاثة ملفات مقلوبة، واحد لكل لغة، مما يستوجب ترجمة المصطلحات من لغة إلى أخرى. لاسيما وأن آلة الترجمة قد تطورت بصورة كافية مما أدى على الأقل إلى القضاء على الفكرة السائدة بأن الترجمة الإنجليزية/ الفرنسية عمل غير مشجع. ولكن المحافظة على الملفات الثلاثة المقلوبة، ومواصلة ترجمة المصطلحات من لغة إلى أخرى داخل نظام يستقبل مدخلات عادية من كل المراكز المضيفة لا يعتبر شيئاً بديهياً.

على أنه لا تزال هناك صعوبات يجب مواجهتها عند البحث فى قاعدة بيانات متعددة المجالات باللغة الطبيعية، حتى فى حالة عدم وجود مشكلة تعدد اللغات. والأكثر بداهة هو أن الباحث يجب أن يثق فى مصادره ليستخرج كل المصطلحات المطلوبة لتحقيق بحث منطقى كامل فى موضوع ما. والحل الأمثل لهذا يتمثل فى بناء قاعدة كبرى للبيانات تشتمل على «جداول من المفاهيم» تخزن داخل الشبكة. ويحتوى كل جدول على كلمات أو جمل إشارية تدل على موضوعه. وعلى سبيل المثال فإن الجدول بعنوان «Financial Conditions» يجب أن يحتوى على العبارة «Finances»، «Profit»، «Loss»، «Balance Sheet» وغير ذلك. ويجب على الباحث أن تكون له القدرة على إدماج جدول كامل ضمن استراتيجية البحث، وربما يكون ذلك باستخدام رقم تعريفى (Knapp) (١). ومن حيث المبدأ فلا مبرر لكى يكون جدول المفاهيم من هذا النوع متعدد اللغات.

وبناء هذا النوع من المكانز ومتابعتها يجب أن يشكلا الوظيفة الرئيسية بالنسبة إلى فريق الضبط المصطلحى الذى يعمل داخل الشبكة. ويظهر جلياً أن قاعدة البيانات المبنية على قوائم المفاهيم لايمكن بأية حال أن تكون نهائية وذلك لأنه من المستحيل أن نتنبأ بالمفاهيم التى سوف يحتاج إليها باحثو الشبكة فى المستقبل. ولهذا يقوم فريق الضبط المصطلحى داخل الشبكة المبنية على اللغة الطبيعية بعمل مستمر يتمثل فى إعداد قوائم جديدة قد تلبى احتياجات المستفيدين الخاصة، وذلك فى تحديث القوائم الموجودة (كإضافة بعض المصطلحات الجديدة).

٢ . ٢ ـ طريقة المكانز الكبيرة/ المكانز الصغيرة

(Macrothesaurus/Microthesaurus Approach)

إن إرساء هذه الطريقة يتطلب تطوير المكنز الحالى لمركز التوثيق والمعلومات بالجامعة العربية حتى يصبح «المكنز الكبير» للشبكة. وهذا يتطلب توسيعاً كبيراً فى بعض المواضيع والمجالات، وربما أيضاً إضافة بعض المجموعات المصطلحية والعلاقات الهرمية الجديدة. ويستمر كل من المراكز الموجودة فى استخدام لغته المقيدة. ويجب أن تدرج كل المصطلحات المتوافرة فى كل المراكز داخل البنية الهرمية للمكنز. وهذا يعنى فى بعض الحالات أن مفاهيم المصطلحات المتساوية مع مصطلحات مركز التوثيق لجامعة الدول العربية، يجب أن تقدم فى شكل إحالة «استخدم». كما يجب معالجة مشكلة تعدد اللغات بنفس الطريقة (أى بربط المصطلحات المتساوية المتعددة اللغات بإجابة «استخدم»). ويجب تشجيع المراكز الجديدة على تطوير مكانز تكون فعلاً مصغرة بالنسبة للمكنز الكبير لمركز التوثيق لجامعة الدول العربية، أى أنها تنصهر فى البنية الهرمية للمكنز الكبير.

وبهذه الطريقة يمكن توحيد كل المدخلات فى شكل قاعدة واحدة للبيانات تحتوى على ملف واحد مقلوب. كما يمكن استخدام المكنز الكبير كوسيلة مساعدة فى عملية البحث داخل القاعدة الموحدة. ولكن هذا فى حد ذاته لا يحل مشكلة توحيد مراكز المعلومات بالشكل الذى يجعلنا نستغنى عن استخدام اللغات المقيدة.

٣ ـ طريقة التحول (Switching Approach)

إن الخيار البديل لقاعدة البيانات الموحدة هو أن تكون الشبكة بشكل يضمن لكل مركز المحافظة على قاعدة بياناته الخاصة داخل حاسوبه الخاص. ويقع ربط كل المراكز داخل شبكة اتصالات مما يسمح لكل مستفيد فى أى مركز بالنفاذ إلى قواعد بيانات كل المراكز داخل الشبكة.

وهذا الأمر يخلق مشاكل عديدة كنا قد تحاشيناها فى حالة قاعدة البيانات الموحدة. ففى المقام الأول يجب أن يوجه المستفيد بطريقة ما إلى قاعدة البيانات التى توفر القدر الأكبر من الاحتمالات التى ترضى احتياجاته من المعلومات. وهذا يمكن حله بعدة طرق مختلفة، لكنه يستوجب أساساً استخدام «منفذ» ما. ويمكن عرض عدة اختيارات على المستفيد للتوجه نحو قاعدة البيانات الخاصة به. ونموذج هذه المنافذ هو «المنفذ الرئيسى للمعلومات» الذى أعده لاتحاد الغربى بالولايات المتحدة: (Western Union In The United States). وبالتناوب، يمكن للمستفيد إدخال موضوع بحثه ومضاهاته «بكشاف قاعدة البيانات» مثل أولائك الذين نجدهم عند أغلب باعة وسائل النفاذ بواسطة الخط المباشر بالولايات المتحدة (مثل كشاف قاعدة بيانات «ديالوغ» (Dialindex) وتتمثل وظيفة الكشاف فى ترتيب البيانات المتوافرة داخل الشبكة، قصد استرجاع الوحدات التى تتلائم وموضوع بحث المستفيد.

وعند القيام بهذا يكون المستفيد من الشبكة قد توجه نحو المصدر الملائم لكن تبقى مشكلتان هامتان:

١ ـ أن برامج البحث المستخدمة فى مراكز المعلومات التى يتجه إليها المستفيد يمكن غير مألوفة بالنسبة إليه.

٢ ـ أن قواعد البيانات التى ربما وقع تكشيفها بواسطة لغة مقيدة، قد تكون أن تكون غير مألوفة بالنسبة للمستفيد.

إن المشكلة الأولى غير بسيطة، وذلك لما تتطلبه

من تطوير لبعض لغات التحكم الموحدة ومناقشة هذا الموضوع تخرج عن نطاق هذه الورقة أما الحلول الممكنة لمعالجة المشكلة الثانية، والتى هى غير بسيطة أيضاً، فسوف نتعرض إليها لاحقاً.

٣ - ١ طريقة توحيد المصطلحات:

(Integrated Vocabulary Approach)

تتطلب معالجة هذه المشكلة توحيد مصطلحات كل قواعد البيانات فى الشبكة وبناء وسيلة بحث يمكن النفاذ إليها بواسطة الخط المباشر وتظهر مختلف المصطلحات المتساوية ضمن هذه المصطحات المتنوعة ويستخدم المستفيد هذه الوسيلة بإدخال مصطلح س. وهذا يفضى إلى عرض يبين كل قواعد البيانات التى يرد فيها س كما يظهر أيضاً المصطلح المساوى لـ س فى قواعد البيانات الأخرى داخل الشبكة ويمكن أن يعتمد التساوى على الترادف، أو على علاقات الجنس والنوع (مثلاً فى إحدى قواعد البيانات المصطلح الأقرب إلى «الألومنيوم» (Aluminium) يمكن أن يكون مصطلحاً أعم: «معدن» (Metals) وهذا النوع من الوسائل المستخدمة لتوحيد المصطلحات يمكن أن يكون ذا جدوى بالنسبة للمستفيد الذى يريد تحويل بحث من قاعدة بيانات إلى أخرى، ويحتاج بذلك إلى معرفة المصطلحات المتساوية الملائمة، لكن هذا الاستخدام يكون ضئيلاً عندما يرغب المستفيد فى إنجاز بحث من مصدر غير مألوف بالنسبة له وبما أن هذه الوسيلة لا توفر هرمية أو أى بنية دلالية أخرى لمساعدة الباحث، فإن هذا يبعث على الشك فى حقيقة جدوى إنشائها، حتى ولو أمكن إعداد كثير من المصطلحات آلياً (Wall and Barnes) (٢) وهذا النوع من الوسائل وقع تطويره واختباره فى الولايات المتحدة من طرف «نيهوف» (وآخرون Niehoff) (٣) - (٧) ذلك أن آخر شكل للنظام الذى اعتمدوه فى تحويل المصطلحات يمكن الباحث من إدخال مصطلح لإنجاز عرض يبين أقرب المصطلحات المتساوية إلى المصطلح المعنى فى كل قواعد البيانات التى يحتوى عليها النظام إن المصطلح الأقرب يمكن أن يعنى مصطلحاً مضاهياً دقيقاً، أو كلمة بديلة، أو مصطلحاً يظهر فى أحد المكانز «كمصطلح مترابط»، أو مصطلحاً محتوى للمصطلح الذى أدخله الباحث:

٣-٢ المعاجم الوسيطة:

Intermediate Lexicon

إن الفكرة وراء ما يسمى «بالمعاجم الوسيطة» هى بناء لغة تحويل حقيقية، لإنجاز وظيفة داخل شبكة مصطلحات شبيهة بتلك التى تنجز من طرف لوحة المفاتيح فى نظام الهاتف وهذا يمكن الباحث من إدخال مصطلح مألوف بالنسبة إليه، ومستخرج من مصطلحات مركز المعلومات المألوف بالنسبة إليه أيضاً.

وهذا المصطلح يمكن تبديله بمصطلح مساو لمصطلحات أى مركز آخر من خلال المعجم الوسيط. والمعجم الوسيط هو فى الحقيقة لغة تحول محائدة. أن كل المصطلحات المستخدمة داخل الشبكة يجب إن تكون «مدونة» فى لغة التحول وعلى سبيل المثال يمكن للمستفيد أن يدخل مصطلح Journals، مما يؤدى إلى رمز قل 952 20 فى المعجم الوسيط، وهذا الرمز يؤدى بدوره إلى مصطلح Periodicals فى قاعدة بيانات خاصة وقع اختيارها من طرف المستفيد. ولتطبيق هذه الطريقة يجب أن تكون مصطلحات كل مراكز الشبكة مدونة فى المعجم الوسيط. ومن الناحية النظرية يمكن استخدام طريق المعجم الوسيط لتبادل السجلات بين قواعد البيانات، أو تمكين مستفيد

الخط المباشر من تحويل استراتيجية البحث من قاعدة البيانات إلى أخرى. أما من الناحية التطبيقية فلم يقع اختيار طريقة المعجم الوسيط إلا فى مجالات قليلة جداً (مثل علم المكتبات)، وليس من الأكيد بأنه يمكن تطبيقها فى شبكة متعددة المجالات ومن مزاياها أنها تعتبر لغة حرة وفى صورة نجاحها يمكن أن تكون قادرة على تحقيق المواءمة بين المصطلحات العربية والفرنسية والإنجليزية أو أية لغة أخرى وإذا كان المعجم الوسيط مرغوباً فيه من طرف «أريسنات»، فإنه من المعقول أن ينطلق فى بنائه من وسيلة موجودة، وإن المرشح البديهى هو مكنز مركز التوثيق لجامعة الدول العربية وفعلاً فإن هذه الطريقة لا تختلف جوهرياً عن الطريقة التى كنا قد ناقشناها فى الفقرة ٢-٢.

الخاتمة:

توجد مشاكل هامة عند توحيد عدة مراكز معلومات مستقلة داخل شبكة واحدة تعمل بطريقة اقتسام المصادر. وهذه المشاكل معقدة بالنسبة إلى «أريسنات» لاسيما وإن هذه الشبكة تتطلب المقابل مع عدد كبير من المواضيع، وأكثر من لغة واحدة وبهذا تكون مشكلة الضبط المصطلحى داخل الشبكة ليست بالسهلة ولكنها فى الوقت نفسه ليست بالضرورة مستحيلة.

٥ ـ مسائل للنقاش:

ـ ما هى الإيجابيات والسلبيات فى عملية توحيد السجلات البيبليوغرافية فى شبكة واحدة داخل قاعدة البيانات فى مقابل الاحتفاظ بقواعد منفصلة وجعل المستفيدين يتحولون من قاعدة إلى أخرى داخل الشبكة؟

ـ ما هى الإيجابيات والسلبيات فى استعمال لغات متعددة (الإنجليزية والفرنسية والعربية) داخل الشبكة مقابل مطالبة المستفدين باستخدام لغة واحدة؟

ـ ما هى الإيجابيات والسلبيات فى بذل جهد كبير يتعلق بإعداد مدخلات الشبكة (مثال ذلك: تطوير معجم وسيط أو بناء مكنز كبير/مكانز صغيرة علائقية فى مقابل جعل المستفيد يتحمل عبء ذلك (أى أن ينصب الاهتمام على المخرجات)؟

المراجع:

1. Knapp, S. D. "Brs/Term. a Database for Searchers." Online '83 Conference Proceedings. Weston, Conn., Online Inc. 1983, pp. 162 - 166.
2. Wall, E., and J. M. Barnes. Intersystem Compatibility and Convertibility of Subject Vocabularies. Philadelphia, Pa., Auerbach, 1969. (Technical Report 1582-100-TR-5).
3. Niehoff, R. T. "Development of an Integrated Energy Vocabulary and the Possibility for On-line Subject Switching." Journal of the American Society Information Science, 27 (1): 3 - 17, Januaty/ February 1976.
4. -"The Optimization and Use of Autoomated Subject Switching for Better Retrieval." Proceedings of the American Society for Information Science, 17: 397-400, 1980.
5. Niehoff, R. T., et al. the Design and Evaluation of a Vocabulary Switching System for Use in Multi-basc Search Enuvionments. Columbus, Ohio, Battelle Columbus Laboratories, 1980, 1980.
6. Niehoff, R. T., S. Kwasny, and M. Wessells. "Overcoming the Data Base Vocabulary Barrier - A Solution." Online, 3 (4): 43 - 54, Octobcr 1979.
7. Niehoff, R T., and G. Mack. Final Report on Evaluation of the Vocabulary Switching System. Columbus, ohio Battelle Columbus Laboratories, 1984.