

الجزء الأول

نظرية نظم استرجاع المعلومات

مقدمة

كثيرا ما يقال إننا نعيش في «عصر المعلومات»، وفي كل يوم تقريبا نعلم عن تطور جديد في تقانة المعلومات. وإن حاجة الانسان إلى المعلومات تنمو باستمرار، مع تزايد اعتماد مجتمعاتنا على المعلومات لتحقيق التقدم والازدهار. وكلنا نحتاج إلى أن نكون قادرين على الوصول إلى الحقائق، ولكننا نحتاج أيضا إلى أن نصل إلى معلومات عن موضوعات معينة - ليس مجرد الحقائق المجردة (العارية) فقط، ولكن تقويم هذه الحقائق واستيعابها في إطارنا المرجعي الخاص. هل يستطيع الحاسب أن يحل جميع مشكلاتنا، أم أننا لا نزال حاجة إلى الاستعانة بالذكاء البشرى للعمل على وضع الحلول؟ كيف نبدأ في إيجاد المعلومات التي نحتاجها؟ يحاول هذا الكتاب أن يبحث نوع المشكلات التي نواجهها في محاولتنا للحصول على المعلومات التي تلبى احتياجاتنا، وكيف يمكن للحاسب أن يساعد في ذلك، وكيف يمكن للجهد البشرى أن يظل ذا قيمة في تسهيل طريقنا إلى الاكتشاف.

قد يكون من المفيد أن نحاول تعريف بعض المصطلحات بحيث يمكننا أن نرى صلتها بحياتنا اليومية، وبعمل اختصاصيي المعلومات. والتعريفات التالية مبنية على تعريفات Concise Oxford Dictionary^(١) وتعريفات Mocquarie Dic-tionary^(٢).

* المعرفة Knowledge هي ما أعرفه

* المعلومات Information هي ما نعرفه، أي: المعرفة المشتركة

* الاتصال Communication هو نقل أو تبادل . . . المعلومات بواسطة المخاطبة، أو الكتابة، أو العلاقات، أى: نقل المعلومات.

* البيانات data [حرفيا المعطيات] كل الحقائق التى يفترض أن تكون موضوعا للملاحظة المباشرة.

* يضاف إلى ذلك، الوثيقة هى أى شكل مادى للمعرفة المسجلة.

من هذه التعريفات يمكننا أن نرى أن البيانات تتألف من حقائق غير مجهزة؛ وأن المعرفة هى ما يمتلكه الفرد بعد هضم الحقائق ووضعها فى السياق؛ وأن المعلومات هى المعرفة المشتركة من خلال خضوعها للاتصال. أما تقانة المعلومات information technology فهى التجهيزات، أى الأجهزة (العتاد hardware) والبرامج (البرمجيات software) التى تمكننا من تخزين ونقل مقادير هائلة من البيانات بسرعة عالية. وحينما ندون أو نسجل المعرفة، فإنها حينئذ يمكن أن تنقل على مسافة ما من حيث المكان والزمان؛ ولا يتعين علينا أن نكون مع المستعلم وجها لوجه كما فى حالة الاتصال الشفهى. وهذا بدوره يقترح مفهوم مستودع أو مخزن المعلومات المسجلة.

وقبل أن تسجل المعرفة، كان الأفراد هم مستودع المعرفة، أى الجسر الذى انتقلت بواسطته المعرفة بين الأجيال المتعاقبة وبين هؤلاء الذين ولدوا المعلومات الجديدة وهؤلاء الذين احتاجوا إلى استخدامها. وإن كمية المعلومات التى يمكن نقلها عن هذا الطريق محدودة، ولذلك بدأ المجتمع يتحرك إلى الأمام حينما بدأ تسجيل المعلومات بين مختلف الأنواع فى أشكال دائمة نسبيا والتى يمكن أن تقوم بدور البديل عن «الأكبر» بالنسبة للإنسان والمعرفة لاتكون مفيدة بصفة عامة إلا عند ما يتم نقلها؛ ومن خلال تسجيلها فإننا نعمل كل ما فى وسعنا لكى نضمن أنها سوف تكون متاحة بصورة دائمة لكل من يحتاجها، بدلا من اقتصارها على شخص واحد.

وفى وقتنا الحاضر، فإن كمية المعلومات الجديدة التى يتم توليدها أو إنتاجها هى من الضخامة بحيث لا يمكن لأى شخص أن يأمل فى متابعتها، بل حتى فى متابعة جزء يسير منها، والمشكلة التى نواجهها هى أن نضمن أن الأفراد الذين يحتاجون المعلومات يمكنهم أن يحصلوا عليها بأقل قدر من التكلفة (سواء فى الوقت أو المال)، وبدون أن تستغرقهم مقادير كبيرة من المواد غير المتصلة وقد وضع شرلوك هولمز⁽³⁾ المسألة بصورة جيدة:

ينبغي للإنسان أن يجعل عقله الصغير مخزنا بسيطا لكل ما يحتاج إلى استعماله من أفكار (تجهيزات)، والباقي يمكن أن يضعه بعيدا فى ذلك المكان من مكتبته الذى يشتمل على الأشياء الزائدة بحيث يرجع إليه حينما يريد.

وقد احتفظ هو لمز نفسه لا بمكتبته التى تحتوى على الأعمال المطبوعة، ولكنه احتفظ كذلك بكشافة الشخصى، الذى رجع إليه فى مناسبات كثيرة ليكمل معرفته الخاصة. والنقطة المهمة هى أننا لا يجب علينا أن نعرف كل شئ - ولكن يجب علينا أن نعرف كيف نصل إلى المعلومات حينما نحتاجها.

وهكذا، فبدلا من الخزانة الفردية للمعرفة، لدينا الآن الخزائن الجماعية أو المشتركة: المكتبات، وخدمات المعلومات، والحاسبات، وبدلا من الذاكرة الفردية لدينا الذاكرة الجماعية: فهارس المكتبات، والبليوجرافيات، وقواعد بيانات الحاسب. وكما أن الفرد الذى تفشل ذاكرته لا يمكن أن يمرر المعلومات المطلوبة، فكذلك الذاكرة الجماعية غير الكافية سوف تخفق فى تحقيق أغراضها. علينا أن نضمن أن الأدوات التى نعدها تلبى حاجات مستفيدينا. ولذلك فإن من المهم جدا أن نحاول تعريف حاجات مستفيدينا بصورة أقرب ما تكون إلى الواقع، وبصورة خاصة فى ضوء النمو الهائل للمعرفة فى السنوات الأخيرة. وتهتم مهنة المكتبات المتخصصة بالمهارات، سواء البشرية أو الفنية، التى يُحتاج إليها لتخطيط واستخدام الأنظمة التى سوف تحقق أفضل النتائج فى تلبية احتياجات المستفيدين. ولازالت المكتبات هى الخزائن الرئيسية للمعلومات، ولكن الحاسبات أصبحت الآن العامل الأهم بين الأدوات المستخدمة.

نمو المعرفة

حددت دراسة قيمة^(٤) ثلاثة عصور eras للحاجة إلى المعلومات، والتي يمكن أن نضيف إليها عصرا أوليا رابعا، هو العصر الموسوعي (متعدد جوانب الثقافة Polymath). لقد مضى على الانسانية زمن كان الحجم الكلى للمعرفة البشرية صغيرا إلى درجة أن شخصا واحدا يمكن أن يستوعبه. ومع نمو المعرفة، فقد بدأنا نتقل إلى عصر التوجه نحو العلوم discipline - orioented، أو فروع المعرفة، والذي استمر في الحقيقة منذ اختراع الطباعة حتى القرن العشرين. وقد اتسم هذا العصر بتقسيم المعرفة إلى فروع أو علوم disciplines أو أقسام Compartments يفصل بينها حواجز قوية أو ضعيفة، تعكس الطريقة التي كانت تُدرّسُ بها؛ وقد نمت علوم جديدة من خلال تجزئ أو «انشطار Fission» العلوم الموجودة، حيث نمت مظاهر جزئية أو خاصة في الأهمية وتطورت إلي علوم مستقلة. وقد تطور العلم عن الفلسفة باعتباره مجالا للدراسة؛ وتطورت الفيزياء عن العلم؛ وتطورت الكهرباء عن الفيزياء؛ وتطورت الالكترونيات عن الكهرباء وفي كل حالة قدم الموضوع تكبيرا أو تجزئنا للقديم، ولكنه بقي في نطاقه. وقد تطورت معظم الأدوات التقليدية للاسترجاع والمستخدممة في المكتبات في داخل هذا الاطار المرجعي، ولم تبدأ كعصور جديدة، مع حاجات المستفيدين المتغيرة، إلا حينما طورت خطط جديدة مثل تعد وغيره من الأدوات المكتبية الحديثة، وهنا بدأت تظهر علامات جدية على الإجهاد. وسوف ندرس بعض هذه المشكلات بعد.

والعصر الثاني وهو عصر التوجه نحو المشكلة Problem - oriented، بدأت أهميته في ثلاثينات القرن العشرين، وبصفة خاصة في الحرب العالمية الثانية. وقد اتسم هذا العصر بالحاجة إلى حل المشكلات الجزئية، باستخدام أى علم تكون ضرورة له، بصرف النظر عما إذا كانت هذه العلوم «تتنمى إلى بعضها» أم لا. وهناك مثال حديث من مجال الهندسة الصغرى micro - engineering، حيث تكون الأجزاء صغيرة إلى درجة يُحتَاجُ معها إلى ابتكار طرق جديدة لتحريكها. وقد استعار العلماء اليابانيون الطرق التي استخدمت بواسطة الأهداب (الشعر

الدقيق جدا الذى يوجد على الأغشية المخاطية) لتحريك المخاط، لكى يكونوا قادرين على تحريك الوحدات الدقيقة الميكروسكوبية المتضمنة. وأفضل مثل معروف هو الهندسة الوراثية، الذى يتضمن اندماج، أو «التحام Fusion»، العلوم مثل الفيزياء وعلم الحياة، وهى علوم اعتدنا على اعتبارها مستقلة أساسا.

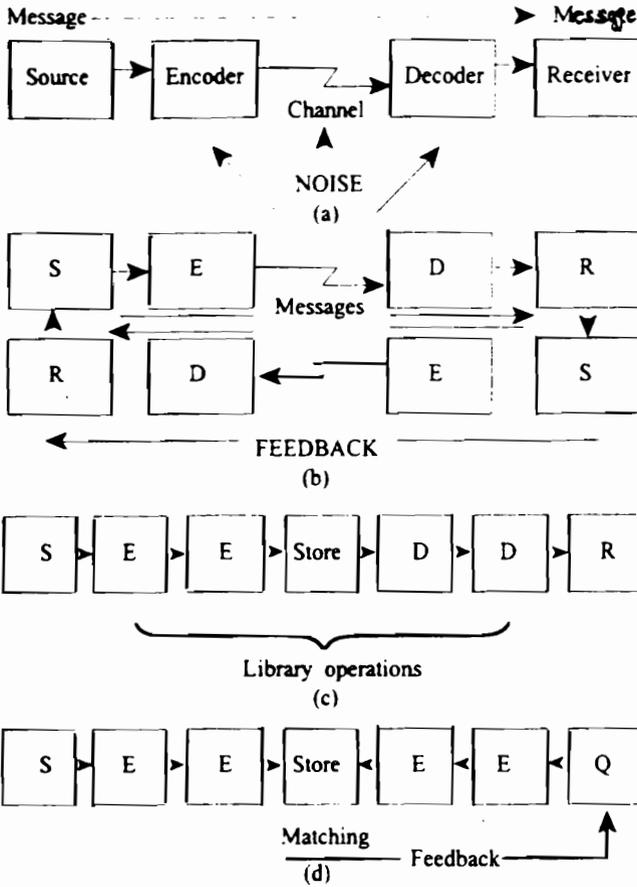
ونحن الآن فى العصر الثالث، وهو التوجه نحو الرسالة mission - oriented ، وفيه قد تمتد الطلبات على المعلومات إلى مدى واسع من العلوم. مثال ذلك: من المؤكد أن طب الفضاء يتطلب معلومات عن الطب، ولكنه فضلا عن ذلك يتضمن مشكلات تنتمى إلى فيزياء الفضاء، والميكانيكا (ظاهرة انعدام الوزن) والنظام الغذائى، الصحة العامة - والقائمة هائلة ومرعبة. ومن الواضح أن الحواجز القديمة بين العلوم، والتي بدأت تزول فى عصر التوجه نحو المشكلة، قد اختفت الآن بصورة مؤثرة، الأمر الذى يقذف بالمزيد من المشكلات فى توصيل المعلومات. وكلما كانت المعلومات الجديدة بعيدة عن مدى المعرفة الموجود لدى الأفراد، كلما أصبح من الصعب بالنسبة لهم أن يستوعبوا وأن يجسدوها فى خزانتهم الخاصة للمعرفة. وإن حاجات المستفيدين للمعلومات فى أيامنا هذه تفرض طلبات على خدمات المعلومات أكثر حدة بكثير من تلك التى جربت فى العصور الماضية، ويجب أن تطور نظم استرجاع المعلومات لدينا بصورة كافية ومناسبة لتلبية هذه الطلبات ويجب أن نميز كذلك بين استرجاع البيانات data واسترجاع المعلومات information. فإن استرجاع المعلومات من الوثائق ليس مثل استرجاع البيانات - وهناك بعض الاختلافات المهمة. إن سؤالاً يطلب بيانات يمكن إجابته مباشرة عن طريق توفير الحقيقة أو الحقائق المرغوبة؛ أما سؤال يطلب معلومات فإن إجابته تتم إما من خلال توفير إحالات إلى الوثائق أو إلى الوثائق نفسها التى من المحتمل أنها سوف تحتوى على المعلومات. إن طلبات البيانات هى إجبارية أو حتمية ولا تحتاج إلى قرارات منطقية من جانب السائل؛ والإجابة التى تقدم إما أن تكون صحيحة أو خاطئة. وعلى الجانب الآخر نجد أن طلبات المعلومات هى احتمالية، وقد تتضمن سلسلة من القرارات المنطقية من جانب السائل. ولذلك فإن طلبات للبيانات

يجب أن تسلم إلى إجابة صحيحة، والإ كانت عديمة الفائدة. وفي مقابل هذا، فإن طلبات المعلومات يجب أن تؤدي إلى إجابة مفيدة، والتي لا يجب أن تكون كاملة بالضرورة؛ وإن فاندتها هي مسألة حكم أو تقدير من جانب السائل.

استرجاع المعلومات كشكل للاتصال

يمكننا أن نرى في ضوء المناقشة السابقة أن أحد مقاييس نجاح نظام استرجاع المعلومات هو فاعليته كوسيلة لتوصيل المعلومات ولذلك فإن من المفيد أن نلقى نظرة على عملية الاتصال نفسها، والطرق التي يمكن من خلالها تعديلها بواسطة عملية استرجاع المعلومات.

يمكننا أن نتناول عمليات استرجاع المعلومات باعتبارها جزءاً من النموذج العام للاتصال. وإن نموذج الاتصال الأكثر شيوعاً هو ذلك الذي ابتكره كل من شانون Shannon وويفر Weaver، والذي يوضحه الشكل (أ) ١,١. وفي هذا النموذج نجد أن المصدر source لديه رسالة message يجب نقلها إلى مستقبل receiver؛ وقبل أن يتسنى نقلها، فإنها يجب أن تشفر للنقل عبر القناة Channel المختارة، وأن تفك الشفرة decoded قبل أن يتسنى فهمها من جانب المستقبل. وفي استرجاع المعلومات فإن المصادر هي منتجو originators الوثائق التي نتداولها؛ وعملية التشفير تتضمن اختيار المظاهر المادية المناسبة - الكلمات، الأصوات، الصور - وترجمتها إلى واسطة مناسبة؛ والقناة هي الوثيقة وانتقالها من المصدر إلى المستقبل؛ وعملية فك الشفرة تتضمن قدرة المستقبل على فهم الرسالة في الشكل الذي قدمت عليه. والعنصر الأخير في النموذج هو الضوضاء noise. ويمكن تعريفها بأنها أي شيء ينتقص من الدقة أو الأمانة في نقل الرسالة من المصدر إلى المستقبل. وقد كان شانون وويفر معنيين بنقل الرسائل عبر أسلاك الهاتف، ولكن مفهوم الضوضاء يمكن تعميمه بحيث يغطي كل أنواع التداخل interference مع الاتصال - مثال ذلك: استرجاع وثائق غير مرغوبة إجابة السؤال ما.



شكل ١,١ نماذج عملية الاتصال

- (أ) نموذج شانون - ويفر
 (ب) الاتصال اللفظي (طريقان)، يتضمن التغذية المراجعة
 (ج) تأثير عمليات المكتبات
 (د) وضع السؤال (الاستفسار): التغذية المراجعة خلال عملية المضاهاة query

وإذا تناولنا الاتصال اللفظي المعتاد، فإننا يمكن أن نرى أن النموذج (شكل (ب) ١,١) هو في الحقيقة الشكل نفسه (أ) ١,١ مضاعفا، بحيث أن المصدر الأصلي يؤدي كذلك وظيفة المستقبل، والعكس. في هذا الوضع، يدخل عنصر مهم آخر إلى الصورة: فكرة التغذية الراجعة Feedback. إذا فسدت الرسالة أثناء انتقالها من المصدر إلى المستقبل، فإن المستقبل يمكنه أن يستفسر فوراً (لم أستطع أن أفهم ذلك تماماً، أو، هل يمكنك أن تشرح ذلك ثانية، من فضلك؟، أو ماذا؟). وهكذا يمكن أن تكون التغذية الراجعة مهمة في التخفيف من آثار الضوضاء.

ولسوء الحظ، فإننا في تعاملنا مع الوثائق، لا نجد إلا تغذية راجعة طفيفة، إن وجدت على الإطلاق، من المستقبل إلى المصدر. فليس للمصدر سيطرة على من يستقبل الرسائل، ومن ثم فلا يمكنه أن يقصرها على مستقبل بعينه. والمستقبلون بدورهم لا يمكنهم أن يكونوا واثقين أنهم فهموا الرسالة بصورة صحيحة، أو أنهم قد حددوا كل - أو أى - الرسائل التي يبحثون عنها. وفي الحقيقة، فإن الوضع يزداد صعوبة من خلال إضافة عمليات تشفير إضافية، والتي سوف تتطلب حينئذ مزيداً من عمليات فك الشفرة من جانب المستقبل. ونحن في المكتبات، اعتدنا على ترتيب الكتب عن طريق وضع رمز Code على الكعب: رقم طلب؛ ثم نحدد ذاتيتها في فهرسنا من خلال سلسلة أخرى من الكودات: مداخل الفهرس (شكل (ج) ١,١). ويمكن لمداخل الفهرس وأرقام التصنيف في الحقيقة أن تعطي قدراً كبيراً من المعلومات، ولكن بالنسبة لمعظم المستخدمين بالمكتبات فإنها تشكل مزيداً من التعقيد في سلسلة الاتصال، وهي مصدر إضافي ثابت للضوضاء. وعلاوة على ذلك، فقد قدمنا تعقيداً آخر: نقل الرسائل يتأخر من خلال وضعها في خزانة من نوع ما. ويمكن أن يعتبر الكتاب خزانة من هذا النوع، ولكن يمكن كذلك أن يعتبر موسوعة على قرص مليزر أو أى نوع آخر من الوثيقة؛ والمكتبات خزائن، وكذلك قواعد البيانات البليوجرافية. وفي هذا السياق فإن مسائل تافهة نسبياً يمكن أن تضيف نصيبها

من الضوضاء فى عملية الاتصال: مثل وضع الكتب على الرفوف فى غير موضعها، أو غير ذلك من العوامل .

أهمية الضبط البليوجرافى

تأخذ الوثائق أشكالاً مادية متنوعة . وإذا أخذنا المكتبة على أنها نموذج مناسب لخزانة الوثائق، فإننا فى وقت ما يمكننا أن نفترض أنه بالنسبة لمكتبات عدة يكون القسم الأكبر من مجموعاتها من الوثائق هو من الكتب؛ فى حين تكون الدوريات أو التقارير الفنية هى السائدة فى مكتبات أخرى . وقد تحتوى المكتبات الآن على المواد السمعية - البصرية والمواد المقروءة بواسطة الحاسب مثلما تشتمل على المواد المطبوعة، وقد تشتمل المواد المقروءة بواسطة الحاسب هى نفسها على بديل للمواد المقروءة والسمعية البصرية فى شكل المواد متعددة الوسائط . وحيث يمكن أن يكون لدينا العمل نفسه فى أشكال مادية مختلفة متنوعة . فتكون لدينا مسرحية هاملت لشكسبير ككتاب مطبوع، أو كفيلم، أو على كاسيت مسموع . وتكون موسوعة ما إما مجموعة من المجلدات المطبوعة أو على قرص مليزر (مدمج) واحد . ويبقى المحتوى الفكرى هو نفسه، والشكل المادى هو الذى يختلف^(٦) . والمهم هنا هو العمل، ولكننا لا يمكن أن نعتمد على الترتيب المادى للوثائق لكى نجمع مع الأشكال المختلفة من العمل الواحد، بل يجب أن نعتمد على بديل - مجموعة من التسجيلات البديلة - للوثائق التى نمتلكها، لكى نحقق الضبط البليوجرافى . وقد لخص هاجلر Hagler أغراض الضبط البليوجرافى:^(٧)

- ١ تحقيق ذاتية وجود كل الوثائق الممكنة فى أى واسطة مادية؛
- ٢ تحقيق ذاتية الأعمال التى تحتوى عليها الوثائق أو كأجزاء منها (مثال ذلك مقالات الدوريات، أبحاث المؤتمرات، محتويات المختارات أو المجموعات الأدبية، إلخ .)؛
- ٣ إنتاج قوائم بهذه الوثائق والأعمال تعد حسب قواعد موحدة للاسناد؛

٤ - توفير كل نقاط الاتاحة المفيدة (الكشافات) إلى هذه القوائم، بما فى ذلك الاتاحة من خلال الاسم، والعنوان، والموضوع على الأقل؛

٥ - توفير طريقة لتحديد مكان نسخة كل وثيقة فى المكتبة أو أى مجموعة يحتاج إلى الوصول إليها.

هذه التسجيلات يجب أن تستخدم الألفاظ؛ فحن لا يمكننا أن نحدد ذاتية المعلومات وأن نحدد مكانها إلا إذا وصفناها وصفا كافيا بواسطة الألفاظ. (وحتى فى مجموعة من المواد الفنية، حيث يكون من الممكن أن نضاهى، وأن ننطق، نماذج مخصصة من textures مع الإيضاحات التى ترد فيها، فسوف يظل من الضرورى أن نحدد ذاتية textures باعتبارها الكلمة المفتاح فى بحثنا).

وقد أصبح الوضع أكثر تعقيدا من خلال الفصل التاريخى بين الضبط الببليوجرافى للكتب والضبط الببليوجرافى للمواد الأخرى. فالكتب تعامل فى العادة على أنها وحدات، وتسجل - تفهرس - على هذا النحو، بينما عوملت الدوريات من حيث الضبط، منذ نشأة الدوريات العلمية فى القرن السابع عشر، وفى مجالات أخرى بعد ذلك، عوملت من خلال آلية مستقلة تماما من المستخلصات والكشافات، لاينتجها المكتبيون فى العادة. ومع نمو المواد السمعية البصرية، نما جهاز جديد من الضبط الببليوجرافى لكى يمكن المستفيدين من أن يجدوا ما يحتاجونه. وقد أمكن التغلب على هذه التقسيمات إلى حدما عن طريق تطوير قواعد بيانات الحاسب، ولكن لازال هناك تقسيم رئيسى بين الضبط الببليوجرافى للكتب والضبط الببليوجرافى للمواد غير الكتب. وحتى فهارس المكتبات على الخط المباشر - الأوباك OPACs، التى نوقشت فى الفصل ١٥، لاتتجاوز عادة هذا التقسيم؛ وبالنسبة للضبط الببليوجرافى للمواد الأخرى فنحن نرجع إلى قواعد بيانات أخرى - رغم أنه يحدث لبس بسبب أن هذه قد تكون متاحة من خلال الأوباك.

ولو كان لدينا نوع واحد من الوثائق فى مجموعتنا، على سبيل المثال مجموعة من التسجيلات فى قاعدة بيانات للحاسب؛ فسوف يتعين علينا أن نصل

إلى التسجيلات من خلال وجهات نظر متعددة. قد نرغب في تحديد ذاتية كل تلك التسجيلات التي تعالج موضوعا معيناً، أو التي ألفها مؤلف واحد، أو التي تحمل عنواناً معيناً، كما أشار هاجلر في فقرته الرابعة آنفاً. ولو كانت قاعدة البيانات تشتمل على الصور، فقد نرغب في تحديد مكان كل التسجيلات التي تحتوى على صورة معينة أو على نوع من الصور. ولكي نجد مواد (وحدات) معينة، فلازلنا بحاجة إلى الضبط البيولوجرافى لكي يعطينا الإتاحة إلى هذه من خلال عوامل متنوعة.

وبعض هذه العوامل يحدد ذاتية المادة أو المواد (الوحدة أو الوحدات) التي تحيل إليها. مثال ذلك: إذا كان لدينا مواصفة قياسية لاختراع، فسوف يكون لدينا مادة أو وحدة واحدة فقط ترتبط بذلك الوصف. وإذا اخذنا مؤلفاً واحداً، فإن عدد الأعمال التي تطابق احتياجاتنا سوف يصبح محدوداً على الفور. وإذا اخترنا عنواناً معيناً، فسوف نجد وحدة واحدة فقط، أو ربما عدداً قليلاً من الوحدات، التي سوف تلبى حاجتنا (المؤلفون يحاولون عادة أن يصوغوا عناوين متفردة لأعمالهم، ولكنهم لا يعرضون عادة الدرجة نفسها من الأصالة!). وعلاوة على ذلك، فهناك الآن معايير موحدة تساعدنا في تحديد الشكل الذى يجب أن نبحث فيه عن المعلومات التي نحتاجها؛ مثال ذلك: قواعد الفهرسة الأنجلو أمريكية الواسعة الانتشار وتسجيلات مارك MARC (نوقشت في الفصل ١٥) تمكنا حالياً من صياغة بحث عن عامل يحدد الذاتية بطريقة تجعله يصبح بحثاً عن البيانات وليس عن المعلومات.

العوامل التي لا تحدد الذاتية

إذا نحن طلبنا معلومات عن موضوع معين، فسوف نواجه مجموعة مختلفة من المشكلات. وأول هذه المشكلات، هي أنه لا يوجد مجموعة من المعايير الموحدة تدلنا بشكل دقيق كيف نعبر عن الموضوع، ويبدو حقيقة أن كل قاعدة بيانات لها معاييرها وقواعدها rules الخاصة. والقراء يطلبون معلومات عن

موضوعات معينة، ويتوقعون من نُظْمِنَا أن تكون قادرة على تقديم الإجابات. وفي هذا الوضع يصبح القراء / المستقبلون مصادر، يفكون شفرة الرسائل في صورة أسئلة. (شكل (د) ١,١). والآن علينا أن نكتشف أية رسائل في خزانتنا تبدو مطابقة لهذه الأسئلة؛ فإذا وجدنا بعضها، فيمكننا أن نمررها إلى السائلين، الذين يمكنهم أن يحددوا إن كانت الإجابات تطابق حاجاتهم. وفي ضوء إجاباتنا، يمكن للسائلين أن يعدلوا من رسائلهم في محاولة لتحقيق مضاهاة أقرب مع احتياجاتهم؛ وبصيغة أخرى، يكون لدينا درجة من التغذية المرتدة في النظام، قد تمكنا في النهاية من إجابة الطلب بالرغم من الإخفاق الأول. هذا الإخفاق قد ينتج عن أسباب متنوعة: قد لا يكون السائلون قادرين على التعبير عن حاجاتهم بوضوح، أو قد لا يكونون واثقين بالضبط ما هذه الحاجات (لو كانوا يعرفون الإجابات ما احتاجوا إلى الأسئلة!)؛ أو قد لا تكون عمليات فك الشفرة كافية أو مناسبة، أو قد لا تكون المصادر الأصلية (المؤلفون) قد جعلوا رسائلهم واضحة، بل حتى قد يكون لديهم رسائل مختلفة في أذهانهم، مثال ذلك: الإجابة عن سؤال من أحد المكتبيين عن العدد الأكبر من الناس المطلوبين للعمل في الإعارة قد تكون موجودة في كتاب عن إدارة السوبر ماركت. والمؤلفون يكتبون في نطاق إطارهم المرجعي الخاص، والذي لن يكون هو الاطار المرجعي نفسه للقراء. وسوف يكون علينا أن نصل بتائجنا إلى الأمثل، في حين أننا نقبل أنه ليس بوسعنا أن نَسِمَ النتيجة بأنها صحيحة أو خاطئة.

الاسترجاع المرجعي، استرجاع الوثائق واسترجاع المعلومات

من المهم أيضا أن ندرك الفرق بين الاسترجاع المرجعي، واسترجاع الوثائق واسترجاع المعلومات. التسجيلات البديلة surrogate records المستعملة للضبط الببليوجرافى، مثل الفهارس التقليدية للمكتبات والببليوجرافيات، تعطينا الاسترجاع المرجعي Reference؛ وسوف نظل مطالبين بأن نقدم الوثائق الفعلية وعلى القارىء حينئذ أن يتفحص الوثائق، ويمكنه أن يصدر حكما نهائيا على

أساس المعلومات التي توجد في الوثائق. وبعض النظم المبنية على الحاسب، مثل النظم في المجال التشريعي، اشتملت على النص الكامل منذ سنوات، وهي هكذا تعطي استرجاع الوثائق في الوقت نفسه مع استرجاع المراجع، ولكن في الوقت الحالي لاتفعل الأغلبية هذا، رغم أن هذا يتغير الآن بسرعة مع تطور الأقراص المليزرة (المدمجة) والوسائط المتعددة. وفضلا عن هذا، فإنه إذا أريد أن تكون عملية الاتصال بين المؤلفين والمستفيد كاملة، فإن المكتبي قد يكون عليه أن يقوم بدور الوسيط. ويفسر الوثائق التي يجدها عن طريق ترجمتها إلى اللغة التي يكن فهمها بواسطة المستفيد. وقد يعنى هذا حرفيا أن يأخذ نصا بالألمانية، مثلا، وينتج ترجمة انجليزية له. وفي كلتا الحاليتين فإنها عملية غالبا ما تهمل، وفي الأعم الأغلب فإن قائمة المراجع تعتبر المنتج النهائي، وليس الاتصال الفعلى للمعلومات.

ويهتم هذا الكتاب بمناقشة مشكلات جعل إجاباتنا على طلبات المعلومات عن الموضوعات أقرب ما تكون إلى الكمال. ولا يعنى هذا أن العوامل التي تحدد الذاتية مثل أسماء المؤلفين لا تطرح أية مشكلات؛ فإن حقيقة أن إنتاج طبعة جديدة من قواعد الفهرسة الأنجلو أمريكية قد استغرق عشرين سنة من المناقشات، والتي روجعت منذ ذلك الحين^(٨)، والتي لازالت موضوعا للمناقشة، هذه الحقيقة تبين بسهولة شديدة أنها تطرح مثل هذه المشكلات! ومع ذلك فإن مشكلات المعالجة الموضوعية للمعلومات هي أكثر حدة لأنها أكثر غموضا؛ ونحن لم نصل إلى المرحلة التي نكون قادرين فيها علي أن نقول أننا قد أنهينا بحثا ما بشكل نهائى. وقد أجرى قدر كبير من البحث حول هذه المشكلات؛ ولا يزال هناك الكثير الذى يجب عمله. وهذا الكتاب محاولة لتوضيح الحالة الراهنة للفن بطريقة تكون مقبولة ككتاب دراسى أولى؛ وهو لا يزعم أنه دراسة متقدمة، والتي يوجد منها الكثير^(٩)، ولكن هدفه أن يعطى المبتدئين شيئا من الفهم للنظريات والأفكار الراهنة.

المراجع

- 1 *The concise Oxford dictionary*, 6th edn, Oxford, Clarendon Press, 1976.
- 2 *The Macquarie dictionary*. St Leonards, NSW, Macquarie Library Pty Ltd, 1981.
- 3 Doyle, Sir A. C., 'The adventure of the five orange pips', *The adventures of Sherlock Holmes*, 1892.
- 4 Arthur D. Little Inc, *Into the information age: a perspective for federal action on information*, Chicago, American Library Association, 1978.
The information society: issues and answers, edited by E. J. Losey, Phoenix, Oryx Press, 1978.
- 5 Blair, D. C., *Language and representation in information retrieval*, New York NY, Elsevier Science Publishers, 1990.
- 6 Hagler, R., *The bibliographic record and information technology*, 2nd edn, Chicago, USA, American Library Association; Ottawa, Canada, Canadian Library Association. c1991.
- 7 Hagler, R., ref. 6 above, 7.
- 8 *Anglo-American cataloguing rules*, 2nd edn, 1988 revision, London, Library Association, 1988.

٩ - لا يبدو أن ثمة نهاية لإنتاج كتب كثيرة عن استرجاع المعلومات والقائمة التالية انتقائية جدا والهدف منها أن تكون مجرد دليل للطلاب وكثير من الأعمال الأساسية نشرت فيما بين ١٩٦٠ - ١٩٨٠ في أعمال مثل:

the Butterworth's series *Classification and indexing in . . . : science and technology*, by B. C. Vickery, 3rd edn, 1975; *the social sciences*, by D. J. Foskett, 2nd edn, 1975; *the humanities*, by D. Langridge, 1976. Specific refer-

ences will be made to these works in particular chapters, and a more complete list is included in the 4th edition of this work. The following list includes some of the works published since the previous edition.

Austin, D., *PRECIS: a manual of concept analysis and subject indexing*, 2nd edn, London, British Library, 1984.

Classification of library materials: current and future potential for providing access, Bengtson, B. G. and Hill, J. S. (eds.), Neal-Schuman, c1990.

Coates, E. J., *Subject catalogues: headings and structure*, reissued with new preface, London, Library Association, 1988.

Craven, T. C., *String indexing*, Orlando, Academic Press, 1986.

Hunter, E. J., *Classification made simple*, Aldershot, England; Brookfield, USA, Gower, c1988.

Lancaster, F. W. and Warner, A. J., *Information retrieval today*, Arlington, VA, Information Resources Press, 1993.

Langridge, D., *Subject analysis: principles and procedures*, London; New York, Bowker-Saur, 1989.

Milstead, J. L., *Subject access systems: alternatives in design*, Orlando, Academic Press, 1984.

Rowley, J. E., *Organising knowledge: an introduction to information retrieval*, 2nd edn, Aldershot, Ashgate, 1992.

Salton, G. and McGill, M. J., *Introduction to modern information retrieval*, New York, McGraw-Hill, c1983.

Subject access: report of a meeting sponsored by the Council on Library Resources Inc, Dublin, Ohio, 1982.

Dym, E. D. (ed.), *Subject and information analysis*, New York, M. Dekker, c1985.

Berman, S. (ed.), *Subject cataloging: critiques and innovations*, New York, Haworth Press, c1984.

Subject indexing: principles and practices in the 90's: IFLA satellite meeting August 17-18 1993, Lisbon, Munich, Saur, 1995.

Turner, C., *The basics of organizing information*, London, Bingley, 1985.

Chan, L.M., Richmond, P. A., Svenonius, E. (eds.), *Theory of subject analysis: a sourcebook*, Littleton, CO, Libraries Unlimited, 1985. A valuable collection of significant articles, referred to elsewhere as *Theory of subject analysis* . . .

Wynar, B. S. *Introduction to cataloging and classification*, 7th edn, by A. G. Taylor, Littleton, CO, Libraries Unlimited, 1985. *1991*

For developments over the past few years it is helpful to consult the chapters on classification in *British librarianship and information work*, London, Library Association, 1982-; 1976-1980, Taylor, L. J. (ed.), 1983; 1981-1985, Bromley, D. W. and Allott, A. M. (eds.), 1988; 1986-1990, Bromley, D. W. and Allott, A. M. (eds.), 1992; these also have comprehensive bibliographies. Useful series of articles include 'Subject access literature' annually in *Library resources and technical services*. Relevant chapters in the *Annual review of information science and technology* are also valuable for the student wishing to pursue the subject in depth.

خصائص نظام استرجاع المعلومات

ينتج المؤلفون كميات كبيرة من المعلومات كل يوم. والتقديرات التى تمت منذ ٣٠ سنة أو أكثر أشارت إلى أن عدد مقالات الدوريات المفيدة (أى التى ليست تكرارية) التى تنشر كل سنة فى العلوم والتقانة وحدهما كان يزيد على مليون^(١)، ولاشك أن هذا العدد قد زاد منذ ذلك الحين. وفى بريطانيا وحدها ينشر ما يزيد على ٥٠٠٠٠ كتابا كل سنة، وقد تفوقت الولايات المتحدة الأمريكية على بريطانيا كأكبر ناشر للكتب فى العالم. وتحتاج المكتبات إلى أن تختار من بين هذه المخرجات الهائلة للاستخدام المباشر لقراءها، ومن خلال الخطط المختلفة للتعاون بين المكتبات يمكن أن يتاح لها اختيارات أوسع بكثير. وفى الطرف الآخر من سلسلة الاتصال نجد القراء، وكل منهم له حاجاته الفردية الخاصة للمعلومات التى يجب أن تختار من بين الكميات المتاحة. وقد يكون مدخل القراء هادفا، أى: قد يبحثون عن الإجابة على أسئلة مخصصة، قد تكون تشكلت فى أذهانهم بصورة أكثر أو أقل وضوحا. هذا هو الوضع الذى سوف نتناوله أولا، ولكننا يجب ألا نهمل القراء المستطلعين browsers، الذين يبحثون عن شىء يتوافق مع اهتماماتهم وليس عن إجابات عن أسئلة مخصصة، والذين يكونون أغلبية المستفيدين فى المكتبات العامة.

استرجاع المعلومات واسترجاع الوثائق

يجب أن نميز أيضا بين المعلومات - المعرفة التي سوف توصل - وبين الوسائل المادية التي يتم بواسطتها هذا التوصيل، كما أشرنا في الفصل ١. وقد اعتدنا في الماضي أن نشير إلى استرجاع المعلومات، بينما الذي كان يوصف هو استرجاع الوثائق، وبمعنى آخر فإننا حينما كانت تطلب منا معلومات، كنا نقدم مجموعة من الوثائق التي اعتقدنا أنها تحتوي على المعلومات المبحوث عنها. وإن نجاح بحثنا كان يعتبر حكما ذاتيا، يمكن أن يصدر فقط بواسطة الشخص الذي يقدم السؤال، وفي الحقيقة فإن الكثيرين من المكتبيين قد اعتبروا أنه سوف يكون خارج نطاق مصطلحاتهم المرجعية أن يقوموا بأية محاولة لتقويم الوثائق التي وجدوها إلا إذا كانت المعلومات المبحوث عنها حقائق خالصة - وحتى بعضها لم يكن كذلك؛ هذا التحفظ لم يكن مقصورا على المكتبيين؛ مثال ذلك: أعطت the International critical tables، والتي نشرت ١٩٢٧ - ١٩٣٣، أعطت مجموعة مختارة من القيم لمعظم البيانات المادية، موضحة ما الذي كان يظن عادة أنه «أفضل قيمة»، ولكن مع إعطاء فصل ومقطع لكل قيمة مسجلة حتى يتسنى للمستفيدين أن يصوغوا فكرهم إن هم رغبوا في ذلك ولا تزال الأكاديمية الأمريكية للعلوم US Academy of Sciences تحتفظ بـ: Office of Critical Tables الخاصة بها، رغم أن الكم الشفاف من البيانات المتاحة الآن أدى إلى نشر طبعة جديدة منقحة من ICT.

وقد جرت محاولات متعددة لتألية ذلك الجزء من النظام المتعلق بتسليم الوثائق، ولكن لم تلق أية واحدة من المحاولات قبولا، ويرجع ذلك أساسا إلى التكلفة وإلى المشكلات الفنية. ومع ذلك فإن تطوير وسائل جديدة ل تخزين مقادير ضخمة من المعلومات في الحاسبات قد قاد إلى مدخل جديد للاسترجاع المباشر للمعلومات. والمعلومات الحقائقية عن موضوعات كثيرة متنوعة هي الآن متاحة في أقطار كثيرة من خلال الأشكال المختلفة من نصوص الفيديو

تكتسب Videotext (النص المرئي)، باستخدام مجموعة التليفزيون التي هي الآن جزء أساسى من أى بيت متحضر (كما أشار جورج أروول George Orwell)^(٢). ويتيح كل من الفرص المميز وقرص البيانات المرئية videodisc مقادير كبيرة متساوية من المعلومات، متضمنة الصوت والإيضاحيات فضلا عن النصوص (الوسائط المتعددة)، باستخدام حاسب يوضع على مكتب (حاسب مكتبي) desktop. وتتيح الشبكات نقل المعلومات مباشرة من حاسب إلى حاسب، وبهذا تمكن من تحقيق المكتب اللاورقى Paperless office الذى ناقشناه فى الفصل ٤ - مع أنه يبدو حتى الآن أن المكاتب تستخدم ورقا أكثر حتى مما استخدم فى الماضى. وقد ذكر أن استهلاك الورق فى الولايات المتحدة الأمريكية سوف يزيد بصورة ثابتة حتى العام ٢٠٠٠ على الأقل!

وينبغى علينا ألا نتجاهل، وليس بوسعنا أن نتجاهل هذه التطورات، ولكن يبقى مهما أن ندرك أن أنها فى الحقيقة لا تغير أساسيات عملية الاتصال، رغم أنها قد تغير بعض المظاهر العملية. وهذا الكتاب يعنى بالمشكلات الفكرية المرتبطة بتلك المظاهر من نقل المعلومات التى ينتظر أن تواجه المكتبي (مستخدمين الكلمة فى أوسع معانيها)؛ وهى تظل هى نفسها مهما كانت الوسائل المادية المستخدمة.

الفحص الجارى والبحث الراجع

Current Scanning and retrospective searching

قد يكون قارئنا مهتماً فى الأساس بأن تكون معلوماته عن المطبوعات الجارية فى موضوعه معلومات حديثة. وفى هذه الحالة لابد أن يكون نظامنا للاسترجاع هو الآخر حديثاً؛ ومع ذلك، فلما كانت الوحدات المحال إليها يسهل الحصول عليها عادة، فإن نظامنا يحتاج فقط إلى أن يكون دليلاً بسيطاً؛ وإذا ما بدا أن وحدة من الوحدات تهم القارئ، يستطيع هذا القارئ أن يحصل على الأصل

بدون مشقة كبيرة. ومن جهة أخرى، فإن القارئ قد يحتاج إلى كل ما يمكن الحصول عليه من معلومات بصرف النظر عن التاريخ؛ وفي هذه الحالة فإن قدرًا كبيراً من المادة قد يصعب الحصول عليه وربما يصبح لذلك مكلفاً، وهنا نحتاج إلى التثبت من أنها ستكون ذات نفع قبل أن نحاول متابعة طلب القارئ. وإن نظامنا لاسترجاع المعلومات ينبغي أن يعطينا معلومات كافية عن الوثيقة حتى نكون قادرين على تقرير أنستمر في متابعتها أم لا. ولما كان الوضع الثاني هو الوضع الأكثر طلباً، فسوف يكون الوضع الذي نركز عليه في هذا الكتاب، ولكننا لن ننسى الفحص الجارى والمستمر. وإن المقارنة بين المعالجتين والطريقتين يوضحها جيداً أعمال مثل:

Current papers in electrical and electronic engineering

Electrical and electronics abstracts

وكل منهما يغطي المجموعات نفسها من الوثائق ولكن الغرض مختلف في كل منهما. وهناك عدد من المطبوعات المتشابهة تغطي مجالات موضوعية متنوعة؛ وفي نطاق المكتبة يمكن مقابلة حاجات الفحص الجارى بواسطة قوائم الإضافات الجارية، في حين يقوم الفهرس بالخدمة الكبرى وهى أنه أداة البحث الراجع فى رصيد المكتبة كله.

البث الإنتقائى للمعلومات

بالإضافة إلى تقديم الوسائل والتسهيلات للفحص الجارى والبحث الراجع وكل منهما يعنى أن المبادرة تأتى من القارئ، فقد أخذت المكتبات منذ عدة سنوات زمام المبادرة بنفسها عن طريق محاولة معرفة أن القراء يعلمون جيداً المواد الجديدة فى ميادين إهتمامهم. وقد يأخذ هذا فى المكتبة العامة صورة عشوائية، أما فى المكتبة المتخصصة، فقد اعتبر هذا العمل دائماً جزءاً من وظيفة المكتبة. ومع ذلك فإن هناك صعوبات معينة فى طريقة إدارة مثل هذه الخدمة بنجاح، بعضها ذهنى وبعضها كتابى. وإن استعمال الحاسب الإلكترونى

يمكن أن يحل كثيراً من هذه المشكلات ويمكننا من إعطاء خدمة أكثر كمالاً ودقة لقرائنا.

وقد طور هـ. ب. لون HP Luhn نظاماً لتشغيل الحاسب الإلكتروني لشركة أ ب م ولازال حتى الآن، زغم أنه قد عدل في بعض الجوانب. ويتضمن هذا النظام في الواقع أن يقوم كل قارئ بصياغة حاجاته بطريقة الوصف الموضوعي نفسها كما تستعمل في تكشيف مجموعات المكتبة. فإذا كانت المكتبة تستخدم قائمة ألفاظ، فإن الألفاظ حينئذ سوف تختار من هذه القائمة؛ وإذا كانت تستخدم خطة تصنيف، تستخدم الخطة لهذا الغرض وتغذى بيانات Profiles القراء هذه في الحاسب الإلكتروني مع البيانات Profiles المشابهة للإضافات الجديدة؛ فإذا وجد الحاسب توافقاً بين الاثنين يطبع لنا إشعاراً بذلك، أو قد يستخدم البريد الإلكتروني لكي يغطي القارئ خدمة مباشرة أسرع.

وعلى هذا النحو تحل المشكلات الكتابية بصورة سهلة نسبياً. أما المشكلات الذهنية فإن معالجتها وحلها أصعب من هذا بكثير. وفي مشروع بحث، ظهر أن المشكلة الضاغطة الأولى في إعداد نظام الـ SDI (بام) Selective dissemination of information يمكن تطبيقه ربما كانت كيفية الحصول على صياغة صحيحة لحاجات القراء. وفي هذا البحث، طلب إلى المستفيدين أن يصوغوا بيانات بمجالات إهتمامهم، وأرسل إليهم مقالات مختارة على أساس هذا. وفي نهاية الشهر طلب إليهم أن يبينوا ما المقالات التي استفيد بها، وأي المقالات التي قرأوا خلال الشهر ثبت أنها مهمة أكثر بالنسبة لهم. وقد اتضح من الإجابات أنه في حين أن غالبية المراجع التي أبلغنا بها أنظمة بام كان لها شئ من القيمة، فإن أهم المقالات، قد وجد أنها ليس لها إلا صلة ضعيفة غالباً ببيانات القراء! وعن طريق سؤال القراء أن يعيدوا الإشعارات، متضمنة مايدل على أن المرجع كان مفيداً أم لا، أمكن إيجاد درجة من التغذية الاسترجاعية استفيد بها في تعديل بياناتهم، ومع

هذا فلن توجد أبدا طريقة يمكن بواسطتها أن نعرف سلفاً المقالة التي يمكن أن تكون ذات قيمة^(٣).

ورغم الصعوبات، فقد طور م م ك IEE عمله في صورة نظام متكامل ومرص، هو إنسبك INSPEC، والذي إندمج معه كل العمليات التي تتضمنها خدمة بام ويتضمنها إنتاج مختلف أجزاء Science abstracts، التي تتضمن:

Electrical and electronics abstracts، و . . Current papers وليست هذه إلا إحدى الخدمات الكثيرة من هذا النوع.

وفي حين أن أنظمة بام ليست قادرة على صنع المستحيل أو تحقيق المعجزات، فإنها يمكن أن تؤدي وظيفتها بفاعلية تامة في داخل مؤسسة معينة، وإن تشغيل الحاسب الإلكتروني يمكننا من توصيل الفوائد إلى جمهور أكبر. وقد أوضح النجاح الذي حققه كثير من الخدمات المتاحة الآن فعلا أنه إذا سلمنا بأن المستفيدين سوف يؤديون ما عليهم ويصوغوا حاجاتهم بدقة، فإن من الممكن تقديم خدمة فعالة جداً على المستوى الوطني.

ونلخص: القراء سوف يحتاجون كل المعلومات التي يمكن جمعها (هذا هو حلم المؤلفين على الأقل!)، ولكن لا يمكن أن نعرف سلفاً أى وحدات المعلومات هي التي ستكون ذات فائدة للقارئ الفرد ومن ثم نطلبها. وما ينبغي علينا عمله هو أن ننظم مكتبتنا بطريقة تجعل من الممكن حينما نبحث عن المعلومات لقارئ ما ألا نضطر إلى تفحص كل محتويات المكتبة لكي نجد ما يريد هذا القارئ، ولكن يكون بوسعنا أن نذهب بأقل قدر من التأخير إلى تلك الوحدات التي ستكون مفيدة. وإذا نظرنا إلى الأمر من زاوية أخرى، نقول أن تنظيمنا بحب أن يسمح باستبعاد ما لا نريده. هذه الفكرة تقدم ثلاثة مفاهيم هامة جداً: الاستدعاء، الصلة، والتحقيق.

الاستدعاء، الصلة، والتحقيق

بالنسبة لأي قارئ يأتي إلى المكتبة محتاجاً إلى المعلومات، ستكون هناك وحدات معينة في مجموعتنا تتصل بموضوع بحثه. ويمكن أن نرتب هذه الوحدات فيما بينها ترتيباً أسبقية؛ فالبعض منها سيكون متصلاً بالبحث على نحو محدد، والبعض الآخر سيكون مفيداً، في حين أن الباقي سيكون متصلاً إتصلاً هامشياً فقط. ولنأخذ على ذلك مثالا؛ قارئ يريد معلومات عن القلط السيامية: قد تكون في مجموعتنا وحدات تعالج القلط السيامية على وجه التخصص، وهذه هي الوحدات المتصلة إتصلاً وثيقاً. وهناك على أية حال عوامل أخرى غير الموضوع سوف تؤثر في ذلك؛ فقد تكون هذه الوحدات مفصلة أكثر من اللازم، أو أقل تفصيلاً مما يجب؛ وقد تكون كتبت دون المستوى العلمي المطلوب، أو في لغة لا يفهما القارئ. وإن خلفية القارئ سوف تؤثر لا محالة على قراره في اختيار الوحدات التي يجدها أكثر إتصلاً. ولكي نجد مزيداً من المعلومات، فقد نوسع من بحثنا: أي نقدم إلى قارئنا تلك الوحدات التي قد لا تعالج موضوعه على وجه التخصص ولكنها تشتمل عليه كجزء من موضوع أوسع. وفي مثالنا، قد نجد وحدات تعالج القلط عامة، وليس القلط السيامية وحدها؛ أو تعالج الحيوانات الأليفة المدللة بصفة عامة، لا القلط وحدها. على أي حال، علينا أن نقبل أنه كلما وسعنا موضوع بحثنا - أي كلما استدعينا مادة أكثر - كلما كان الأرجح ألا تكون المواد متصلة. وفي الدرجة القصوى سوف نجد معلومات متصلة في دائرة معارف عامة، ولكنها سوف تكون جزءاً صغيراً جداً من المعلومات الكلية التي تحتوي عليها. وعلى ذلك فإن ثمة علاقة عكسية بين الاستدعاء recall - عدد الوحدات الإضافية التي نجدها مع توسيع بحثنا - والصلة relevance - احتمال توافقها مع احتياجات قرائنا.

والقارئ يكتفي عادة بوحدات قليلة، طالما أنها تحتوي على نوع المعلومات التي يريدها؛ أي أننا بحاجة إلى نظام يعطينا درجة عالية من الصلة، حتى ولو

كان الاستدعاء محدوداً. ولكن ثمة أوضاعاً سوف يطلب فيها القارئ إستدعاء عالياً - أكبر قدر ممكن من المعلومات - حتى وإن كان هذا يعني أنه سوف يبحث وسط عدد كبير من الوحدات التي سوف يتضح بالبحث أنها محدودة أو حتى معدومة القيمة بالنسبة له. فنحن بحاجة إلى أن يكون باستطاعتنا تنويع إجابات نظامنا لكي يتوافق مع نوع الطلب. ومن الواضح أيضاً أن الصلة عامل ذاتي يعتمد على الفرد؛ فالسؤال الواحد يطرحه قارئان مختلفان، قد يتطلب إجابتين مختلفتين. ويمكن في الحقيقة أن نمضى في المناقشة أبعد من ذلك. فكل وثيقة كشف عنها بحثنا قد تغير رأى القارئ فيما هو متصل، كما ذكرنا في الفصل ١، وهكذا فإنه حتى القرارات الفردية للشخص الواحد قد تختلف من حين لآخر.

وتنشأ المشكلة من حقيقة أن القراء يبحثون عن معلومات يكونون بها حاصيلة معارفهم - إطارهم المرجعي - بأقل قدر من الصعوبة، في حين أن المؤلفين يقدمون المعلومات في شكل يمليه إطارهم المرجعي، وكل مناله إطاره المرجعي، ولذلك فلن يتطابق الاثنان تطابقاً كاملاً. ويجب علينا أن نصمم نظامنا لاسترجاع المعلومات بحيث تمكن من تحقيق أعلى درجة ممكنة من التوافق مع حاجات القراء، ولكن يتعين علينا أن نقبل أن هذه النظم لن تكون كاملة أبداً.

وقد قاد الرأى المتعلق بالصلة إلى مفهوم وثاقة الصلة *Pertinence* أو المنفعة *utility*. فلو أن وثيقة ما استرجعت إجابة لسؤال معين، فإن درجة صلتها يمكن تقديرها بواسطة مجموعة من هؤلاء المتمرسين في الفن^(٤)، ولكن وثاقة صلتها لا يمكن تقديرها إلا بواسطة صاحب السؤال. وبمعنى آخر، فإن الصلة هي حكم إجماعى، ووثاقة الصلة حكم فردى. ويمكن أن ننظر إلى المسألة بطريقة أخرى، وهى أن وثيقة ما استرجعت لإجابة سؤال ما قد تكون مفيدة للسائل، ولكن فائدتها (منفعتها) قد تتغير؛ مثال ذلك: إذا استرجعنا الوثيقة نفسها فى بحث ثان، فسوف تكون فقدت منفعتها فى المرة الثانية *round*. الصلة لن تتغير، ولكن رأى السائل سوف يتغير.

ومن الممكن فى وضع تجريبي، كذلك الذى يحدث فى دراسة عن فاعلية نظم التشفير المختلفة، من الممكن أن تصدر الأحكام عن الصلة سلفا، مثال ذلك: عن طريق فحص كل الوثائق فى المجموعة المختبرة بالنسبة لكل الأسئلة الاختبارية، كما حدث فى مشروع كرانفيلد الثانى^(٥). وحينئذ يمكن أن نصل إلى رأى موضوعى عن نجاح النظام من خلال مقارنة النتائج التى تم تحقيقها مع هذا النظام مع الاجابات التى سبق تحديدها. ومن المعتاد فى هذا الوضع أن نشير إلى التحقيق لا إلى الصلة. ويستخدم مصطلح التحقيق precision على نطاق واسع فى الانتاج الفكرى فضلا على مصطلح الصلة، ولكننا فى كتابنا هذا سوف نستخدم مصطلح الصلة حينما يتضمن الأمر حكما ذاتيا. وقد ناقش لانكستر المصطلحات المختلفة فى ثلاث مقالات فى: Encyclopedia of library and information science^(٦) وقد أظهرت دراسة أحدث أن العلاقة العكسية بين الاستدعاء والصلة، والتى عرفت لأول مرة فى مشروع كرانفيلد الأول^(٧)، يمكن إثباتها بطريقة رياضية. وإذا أردنا أن نحصل على استدعاء محسن وصلة محسنة فيجب علينا أن نغير استراتيجيتنا للبحث^(٨).

ويمكن أن نستخدم رسوم فن Venn البيانية ورمز الفئات لكى نفحص هذه المفاهيم بصورة أدق. إذا أخذنا كعالم لنا مجموعة من الوثائق I، فلإجابة على سؤال معين لابد أن يكون ثمة فئة A من الوثائق المتصلة، حيث تكون A فئة فرعية من L ($B \subset L$). فإذا إستخدمنا نظامنا لاسترجاع المعلومات لمحاولة إيجاد هذه الوثائق، فإننا فى الحقيقة سوف نسترجع فئة مختلفة B ($B \subset L$)، والجزء المتصل منها هو الفئة الفرعية التى تكون نقطة التقاطع (التداخل) بين A و" $A \cap B$ ". ولنحاول الآن تعريف المصطلحين نسبة الاستدعاء ونسبة التحقيق:

$$\text{نسبة الاستدعاء} = \frac{(A \cap B)}{A} = \frac{\text{(الوثائق المتصلة المسترجعة)}}{\text{(مجموع الوثائق المتصلة)}}$$

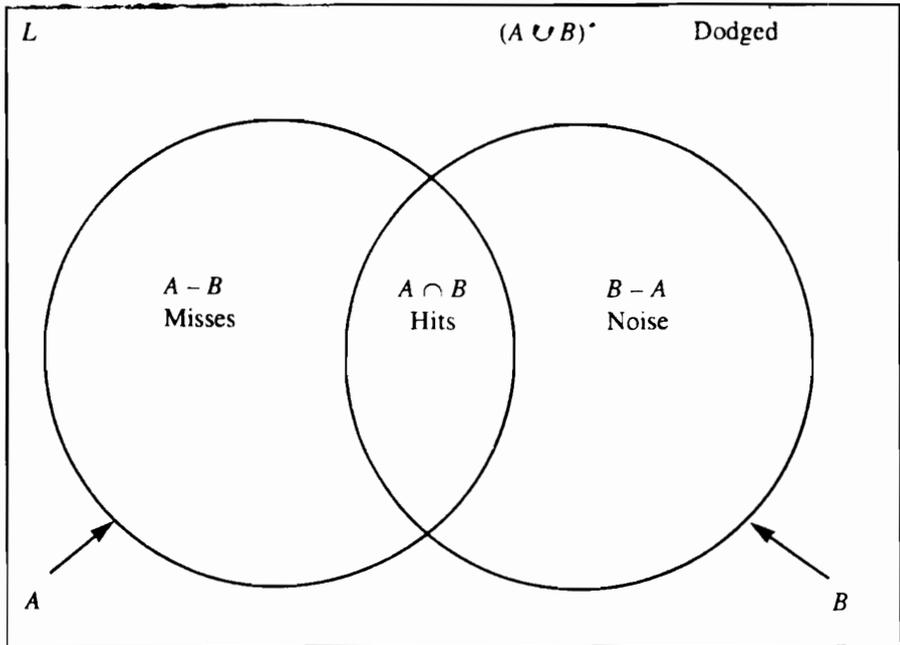
$$\text{نسبة التحقيق} = \frac{(A \cap B)}{B} = \frac{\text{(الوثائق المتصلة المسترجعة)}}{\text{(مجموع الوثائق المسترجعة)}}$$

ويتم التعبير عن هذه عادة على أنها نسب مئوية عن طريق المضاعفة بواسطة ١٠٠. وهناك مصطلح آخر يستخدم أحياناً هو نسبة Fall - out النزول (الهبوط)

ويعرف على النحو الآتي:

$$\text{نسبة النزول} = \frac{(A - B)}{B} = \frac{\text{(الوثائق المتصلة التي لم تسترجع)}}{\text{(كل الوثائق التي لم المسترجع)}}$$

والفئة B - A التي تتألف من الوثائق التي استرجعت ولكنها غير متصلة، يمكن اعتبارها ضوضاء، في حين أن الفئة (A U B)، الوثائق التي لا هي مسترجعة ولا هي متصلة، يمكن اعتبارها مراوغة، في اصطلاح فيكرى.



شكل ٢,١ رسم فن عن عملية الاسترجاع

ومن العسير أن نجد تعريفاً واضحاً للفئة A لسوء الحظ. وربما لانجد لها وصفاً دقيقاً إلا في الأوضاع التجريبية فقط. في الحياة الواقعية هناك منطقة ظلال، تتألف من تلك الوثائق التي قد تكون متصلة. فإذا رسمنا مقطعاً عرضياً في A، وعينا موقعه على رسم بياني يوضح درجة الصلة، فسوف نجد النتيجة التي يدل عليها الاختصار APUPA والذي وضعه رانجاناثان: تعبر U عن Umbra

أى تلك الوثائق التي من الواضح أنها متصلة (فى نطاق ظل الموضوع) تعبر P عن Penumbra «المنطقة غير الواضحة»؛ وتعتبر A عن Alien، أى تلك الوثائق التي من الواضح أنها غير متصلة. وإن المنطقة غير الواضحة Penumbra هي التي تجعل من الممكن تعريف A بوضوح، وهذا يعنى أننا لا نستطيع أن نستخدم لفظ التحقيق فى هذا السياق. وقد استعملنا فى هذا النص لفظ الصلة ليدل على وضع الحياة الواقعية، والتحقيق ليدل على الوضع التجريبي حيث يمكن تحديد A سلفاً.

الفئات المشوشة Fuzzy Sets

لقد كان رانجاناثان بحكم دراسته متخصصاً فى الرياضيات، ولذلك فربما لاندهش إذا وجدنا أنه قد استخدم جزءاً من نظرية الفئات الرياضية فى محاولاته للتحليل الكمي quantity لمفهوم APUPA عنده. وهذه هي فكرة الفئات المشوشة. فى نظرية الفئات القياسية، وإذا كان عالمنا هو L الذى X فئة فرعية منه، يمكننا أن نعرف وظيفة عضوية $FA(X)$ التي ستكون I إذا كانت A عضواً فى A وصفرًا إذا لم تكن؛ بمعنى آخر إما أن تكون X أولاً تكون عضواً فى الفئة A. فإذا كانت A فئة مشوشة، فإن $FA(X)$ قد تأخذ قيمة ما بين صفر و I؛ بمعنى آخر فإن X ليست: $FA(X) = 0$ ؛ بل هي $FA(X) = 1$ ؛ أو قد تكون $(0 < FA(X) < 1)$ ؛ عضو فى A. وقد طبق المفهوم على نطاق واسع منذ نشره الأصيلى بواسطة Zadeh⁽⁹⁾ فى سنة ١٩٦٥، وقد سجل: Social Science citation index عدة فئات من المراجع تستشهد بهذا، لكن ليس من بينها إلا عدد قليل عن استرجاع المعلومات.

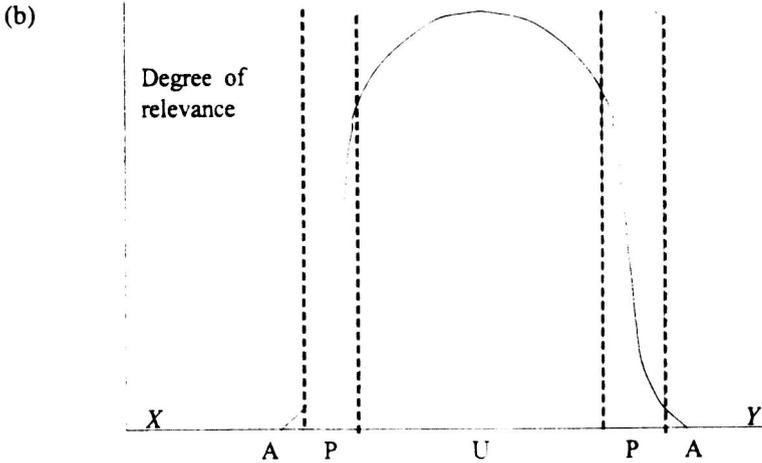
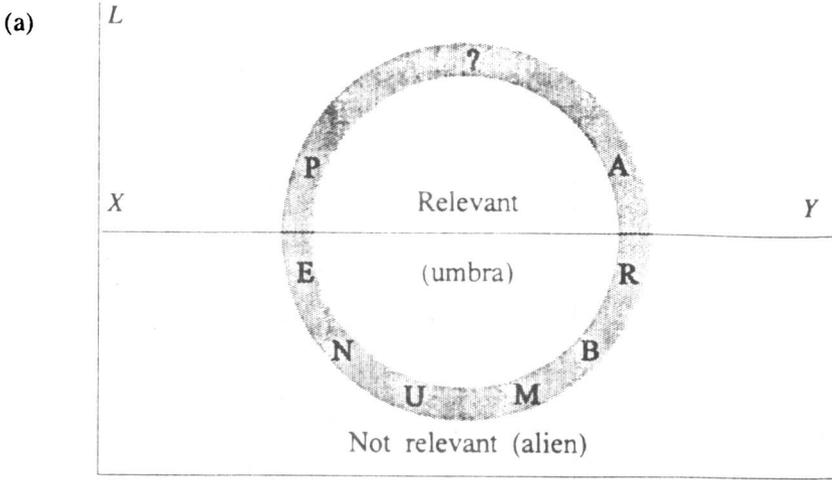
والهدف من تطبيق نظرية الفئة المشوشة على نظم استرجاع المعلومات هو أنه مع استخدام الحاسب، سوف نكون قادرين على ابتكار نظام يتيح لنا أن نرتب الوثائق المسترجعة فى ترتيب محتمل للصلة، بدلا من النتيجة «إما / أو» للبحث التقليدي، والتي تفرز الخراف sheep المتصلة عن الماعز goats غير المتصلة وتتجاهل حقيقة أن هناك كمية كبيرة من «ربما» المتضمنة. والنظم التي ترتب مخرجات الوثائق هي الآن موجودة، وتستخدم وظيفة مضاهاة بين السؤال والوثيقة لترتيب المخرجات. وتستخدم نظم أخرى وظائف الاحتمال للغرض

نفسه . ويبقى أن نعرف هل تقدم نظرية الفئات المشوشة نتائج أفضل . بعض الرياضيين لازالوا متشككين حول قيمة النظرية ككل ، في حين أن كتابا آخرين قد اقترحوا أنها من الممكن أن تطبق على استرجاع المعلومات .

منحنى الاستدعاء - التحقيق

يمكننا أن نعرض الأقسام الأربعة للوثائق:

A, ÇB, A - B, B - A and (A È B)

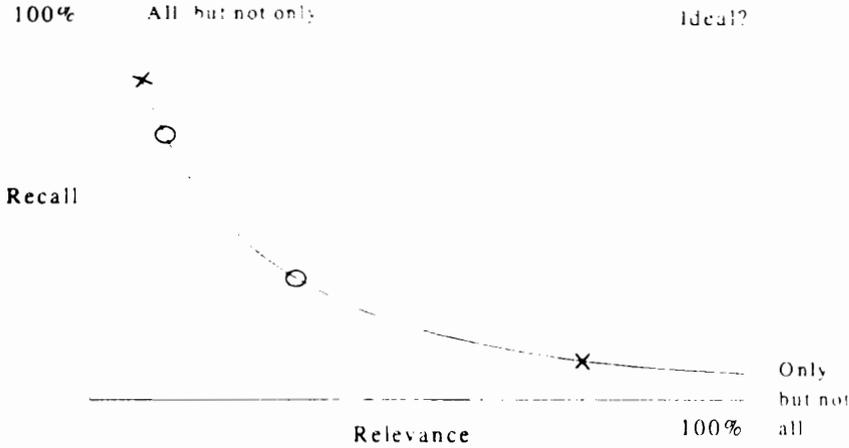


شكل ٢،٢ مشكلة تعريف القسم A من الوثائق المتصلة:
 (a) الحدود بين A و A لا يمكن تحديدها
 (b) APUPA التي وضعها رانجاناثان عن الجزء المتقاطع

في شكل مصفوفة:

| المجموع | غير المسترجعة | المسترجعة | |
|---------|---------------|------------|-------------|
| A | A - B | $A \cap B$ | المتصلة |
| A' | $(A \cup B)'$ | $B - A$ | |
| <hr/> L | <hr/> B' | <hr/> B | غير المتصلة |

من هذه المصفوفة، يمكننا أن نرى لماذا يفضل بعض الكتاب أن يقارنوا نسبة الصلة ونسبة السقوط (النزول) وليس نسبة الصلة ونسبة الاستدعاء. والأخيران لهما البسط نفسه $(A \cap B)$ ، في حين أن نسبة السقوط تستخدم الجزء المكمل من المصفوفة، وهذا يعطى بسطا ومقاما مختلفا عن كل من الآخرين: $(B - A)$ و A' بدلا من $(A \cap B)$ و A أو B (١٠). ومع ذلك فإن الاستدعاء والصلة مألوفان أكثر ويمكننا أن نأخذ نتائج عدد من الاختبارات ونستخدمها نرسم رسما بيانيا لنسبة الاستدعاء مقابل نسبة التحقيق. وسوف يتركز رسما بطبيعة الحال، وبصورة مثالية على زاوية ١٠٠٪ استدعاء و ١٠٠٪ تحقيق، ولكن من الناحية العملية سوف نحصل على منحنى من النوع الذي يوضحه الشكل ٢،٣.



شكل ٢،٣ منحنى الاستدعاء/التحقيق

مرونة النظام: القدرة على

التقاط نقطة التشغيل المناسبة على منحنى الاستدعاء - الصلة

نظام X - X نظام أكثر مرونة من نظام O - O (O و X) تصعد الحدود لبارامترات التشغيل

ودلالة هذا المنحني هي أننا إذا حاولنا تحسين الاستدعاء، فسوف يتم ذلك على حساب التحقيق، والعكس، فإذا أردنا تحسين التحقيق سيكون ذلك على حساب الاستدعاء. ونحن ندين بالألفاظ: الاستدعاء، والصلة، والتحقيق لمشروع كرانفيلد، ولكن الفكرة نفسها قد عبر عنها قبل ذلك بعدة سنوات فيرثورن Fairthorne في العبارتين: 'All but not' (OBNA) 'Only but not all' (ABNO) only: (١١) وإن أحد مقاييس نظام إسترجاع المعلومات هو الحرية التي يمكن أن تنتقل بها من جزء في منحني الاستدعاء - التحقيق إلى آخر؟ مثال ذلك: إذا لم يكشف بحثنا الأول عن كل المعلومات التي نريد، هل نستطيع زيادة فاعلية الاستدعاء بتحريك المنحني (ومن ثم التضحية بدرجة من التحقيق)؟ وإذا كشف بحثنا الأول عن كمية هائلة من المعلومات، هل يمكننا زيادة التحقيق ومن ثم إعادة تشكيل إستراتيجيتنا للبحث لتضييق الاستدعاء؟ وسوف نناقش في حينه الطرق التي تتخذ لتغيير إستراتيجيات البحث لإعطاء مثل هذه التغييرات، حينما يبدو جلياً أنه ليس كل الأنظمة تمتلك الدرجة نفسها من المرونة في هذا السياق.

وينبغي أن نكون حذرين من قبول منحني الاستدعاء - التحقيق دون مناقشة. فكما أشار كليفر دون Cleverdon نفسه، يمثل هذا المنحني متوسط أداء أي نظام، وهذا قد يختلف إختلافاً كبيراً في الأوضاع المخصصة. مثال ذلك: في الدراسة التقويمية التي قام بها لانكاستر (الفصل ٢٨) وجد أن النظام يعمل بمتوسط يقرب من ٥٨٪ إستدعاء و ٥٠٪ تحقيق، ويسترجع في كل بحث متوسطاً من ١٧٥ وثيقة. ولكي نحقق ٨٥٪ إلى ٩٥٪ إستدعاء معناه إسترجاع متوسط بين ٥٠٠ إلى ٦٠٠ وثيقة في كل بحث، بمعدل تحقيق يقرب من ٢٠٪. على أي حال، إذا نحن تفحصنا نتائج البحوث الفردية نجد أنه قد تحقق في بعض الحالات ١٠٠٪ إستدعاء مع ١٠٠٪ تحقيق، بينما في بعضها الآخر كان الاستدعاء والصلة صفرأ! وفضلاً عن ذلك، فإنه في حين قد لا يمكن تغيير إستجابة النظام لسؤال معين - فإننا إذا حاولنا تحسين الاستدعاء أو التحقيق فإن هذا يمكن أن يحدث فقط في أحدهما على حساب الآخر - من الناحية العملية،

حينما نبدأ فى الحصول على الوثائق من النظام إجابة لسؤال فقد نكون قادرين على تعديل السؤال بطريقة تجعل من الممكن تحسين كل من الاستدعاء والتحقيق.

وبالنسبة للمستفيد، فإنه عادة يفضل الصلة وليس الاستدعاء. وإن غالبية الأسئلة يمكن إجابتها عن طريق تقديم مالا يزيد عن إثنتى عشرة وثيقة، على فرض أنها جميعاً مفيدة، وهناك حالات قليلة فقط هى التى تتطلب إجابتها إستدعاء أعداد كبيرة. مثال ذلك: يحتاج الباحث عن براءة إختراع ما أن يتأكد أنه وجد كل وثيقة متصلة، حيث أن النشر السابق هو أساس لابطال براءة الإختراع؛ وإن واحدة كهذه تكون كافية، ولكن النظام يجب أن يصمم بحيث يكون الباحث واثقا من أنه سوف يجد تلك الوثيقة إذا كانت موجودة. ومن ناحية أخرى فإن السائل العرضى لن يحتاج فى العادة إلى الدرجة نفسها من الثقة، ونادرا ما يرغب فى أن يقدم عدد كبير جدا من الوثائق استجابة لطلبه.

وهناك أيضا نقطة أن وثائق كثيرة تكون تكرارية: فهى لاتضيف أى شىء مهم إلى معرفتنا. وقد فحص ر. شو R. Shaw مجموعة من الوثائق عن حشيشة اللبن milkweed بالتفصيل، وخلص إلى أن كل المعلومات التى يمكن أن نجدها تحتوى عليها ٩٦ وثيقة من بين عدد إجمالى بلغ ٤٠٠٠! (١٢) وبطبيعة الحال فقد كان عليه لكى يميز الـ ٩٦ أن يفحص العدد كله أى ٤٠٠٠، ولكن هذا يؤكد نقطتين. الأولى منهما هى أنه حتى الانتاج الفكرى الذى يُزعمُ أنه أصيل قد يكون عملا سابقا؛ والنقطة الثانية هى أهمية العاملين الوسطاء الذين عملهم إجراء عملية تصفية بين المؤلفين والمستفيدين. وهناك مجال كبير لمزيد من التحليل لمحتويات الوثائق، فى مقابل تزويد المستفيد بها دون أى تساؤل الأمر الذى يضع على عاتقه حينئذ أن يفصل القمح عن التبن. وقد قيل أننا نشهد انفجارا فى النشر وليس انفجارا فى المعلومات؛ وربما يستطيع المكتبيون حل بعض مشكلات الاستدعاء والصلة عن طريق تشجيع نوع ما من منع الحمل

الإدبى Literary Contraception !

ومع ذلك فينبغي ألا نغفل حقيقة أن هناك حاجة حقيقية لدرجة معينة من التكرار. فالأفكار الجديدة تحتاج إلى أن تثبت على أكثر من مستوى؛ وما يكون مثاليا بالنسبة لأنداد المؤلف ربما لا يزيد في حكمة وفهم الشخص العادي، الأمر الذي يحتاج إلى شرح على مستوى أبسط. ومع الاتجاه الحديث إلى التخصص المتزايد، قد يكون لدينا حتى مستويات من «الشخص العادي»، الذي يحتاج إلى سلسلة من مستويات الشرح والتفسير تتدرج من المقال الأصلي في مجلة علمية إلى عرض تليفزيوني وتقرير في صحيفة. وهناك أيضا مشكلة النشر بلغات أخرى غير لغة الأصل. وقد تحتاج، وهذا أمر يعتمد على الغرض من نظامنا لاسترجاع المعلومات، إلى خزن المعلومات على أكثر من مستوى، وأن نضمن أن لدينا الوسائل لاختيار المستوى المناسب لأي مستفيد معين.

وقد وضع بليير Blair مفهوما مهما ومفيدا لتوضيح الموقف «المعتاد» من الاستدعاء والصلة، ذلك هو نقطة اللاجدوى Futility point. (١٣) إن نسبة الصلة التي تكون مقبولة تماما في نظام صغير قد تكون غير مقبولة على الإطلاق في نظام مبنى على الحاسب يوفر الإتاحة إلى ملايين المراجع. وقد صُكَّ مصطلحان لوسم رد فعل المستفيد تجاه نتائج بحث ما. ولكي نرضى المستفيد، فإن البحث يجب أن يسترجع مجموعة صغيرة من الوثائق بدرجة تجعل الباحث راغبا في تصفحها لكي يستخرج منها الوثائق المتصلة. وإن عدد الوثائق التي يكون الباحثون راغبين في تصفحها قبل أن يتوقفوا في اشمزاز هو نقطة اللاجدوى (العَبث) Futility point FP. وإذا كانت نلا FP لباحث ما هي n ، فحينئذ يكون الباحث راغبا في نصف n من الوثائق من الفئة B في شكل ١، ٢ قبل أن يتوقف دون أن يجد الوثيقة (الوثائق) التي يريدتها من الفئة A. ومع ذلك فهناك أيضا نقطة لاجدوى متوقعة، m ، وهي حجم الفئة التي يرغب الباحث في أن يبدأ تصفحها أو استطلاعها؛ مثال ذلك: من الواضح أنه في الغالبية العظمى من الحالات لا يؤدي الاعلان عن أن البحث قد استرجع ١٢٠٠ مرجعا، لا يؤدي هذا إلى زيادة حماسة الباحث. ويقترح بليير، بناء على الملاحظة، أنه من

الناحية العملية تكون m من طبقة ٣٠؛ ومع وجود قواعد بيانات كثيرة تحتوى على ملايين المراجع، فإننا قد يتعين علينا أن نعيد النظر فى أفكارنا عن المستويات المقبولة للصلة.

إن التأكيد على مفهوم نقطة اللاجدوى قد أتى من دراسة أجدها لانتز Lantz عن: The London University Computer Information Service LUCIS^(١٤). وقد طلب إلى المستفيدين الإجابة على استبانة تضم أكثر من ٢٠٠٠ سؤالاً؛ وقد طلب إليهم السؤال الأول أن يقدروا عدد المراجع المتصلة المسترجعة، وطلب الثانى عدد المراجع التى قرأوها فعلا. وقد أظهرت الاجابات أنه كلما زاد عدد الوثائق المتصلة التى يتم استرجاعها، كل قل الجزء الذى قرأه المستفيدون فعلا. وحينما يكون العدد المتصل ١٠، يكون العدد المقروء ٥؛ وحينما يكون العدد المتصل ١٠٠ يكون العدد المقروء ٢٨ تقريبا. وقد اختلف الجزء المقروء بحسب المجال الموضوعى؛ بالنسبة للهندسة، وصل العدد المقروء حدا هو ١٠ تقريبا؛ وفى الطب وصل إلى ٦٠ تقريبا. وفى العلوم الاجتماعية، وهو المجال الذى لاحظته بليز، وصل الرقم إلى ما يقارب ٣٠. وقد أظهر التحليل الرياضى أن منحنى العدد المقروء y مقابل العدد المتصل X يمكن أن يطابق تماما التعبير الأسى exponential:

$$y = a(1 - e^{-bx})$$

حيث المعنى القيمى لـ a هو ٣٤,١ ولـ b هو ٠,١٧, ٠,٠٠٠. وحيث أن:

$$x \rightarrow \infty, e^{-bx} \rightarrow 0, \text{ so that } a(1 - e^{-bx}) \rightarrow a$$

إذن يكون معنى الحد الأقصى للرقم المقروء هو ٣٤ تقريبا. وهذه أرقام متوسطات حسابية (أفاد قارئان فى الحقيقة أنهما قرأ أكثر من ٢٠٠ مرجعا) ولكن التنشا به مع رقم كبير إلى حد أنه لا يمكن تجاهله. وتمتلك مكتبة جامعة لندن بطبيعة الحال موارد جيدة جدا لخدمة المستفيدين؛ وفى المكتبات ذات الموارد القليلة، فإن من المحتمل أن الجزء المقروء سوف يتأثر بواسطة عامل

الإتاحة availability . والقراء يفضلون فى العادة المواد المتاحة فى الحال، حتى ولم لو تكن «أفضل ما يوجد»، على المواد التى يتعين الحصول عليها من أماكن أخرى، حيث يتأخر الحصول عليها. وفى المكتبة ذات الموارد القليلة، فإن من المنتظر أن يكون الجزء المقروء فعلا من المواد المسترجعة أقل مما وجد فى هذه الدراسة.

وعلى الرغم من الحاجة إلى معرفة هذه النتائج، فمن الواضح أن الاستدعاء والصلة (التحقيق) هى مفاهيم مفيدة فى دراسة أى نظام لاسترجاع المعلومات. ومع ذلك، فهى ليست المعايير الوحيدة التى يمكن أن يحكم بها على أن نظام. ويمكننا الآن أن نتناول بعض العوامل التى تؤثر على الاستدعاء والصلة، وبعض المظاهر المهمة الأخرى لنظم استرجاع المعلومات.

احتمال الخطأ

المكشوفون بشر، وكذلك المستفيدون، ومثلهم العاملون على لوحة المفاتيح الذين ينتجون النص المقروء آليا. وجميعهم عرضة للخطأ. ويجب أن يكون نظامنا نظاما يقلل من احتمال الخطأ إلى أبعد حد ممكن. وقد أظهر بحث أجراه مهندسو التليفونات من سنوات مضت أن^(١٥) احتمال الخطأ فى الاتصال التليفونى يرتفع بصورة حادة إذا وصل طول الرقم إلى تسعة أعداد أو أكثر. (نظرة على اتجاهات أطوال رقم التليفون يمكن أن تكون محبطة). فإذا كان النظام يستخدم الأرقام للتشفير (التكويد)، مثل نظام التصنيف، فإن الأخطاء سوف تصبح شائعة إذا زاد طول الرمز على الحد المذكور آنفا. وحتى لو استخدمنا الكلمات، فسوف يبقى احتمال الخطأ قائما؛ مثال ذلك: لا بد أن كثيرين من المستفيدين الانجليز من مدلاين MEDLINE مغتاظون بسبب عدم قدرة النظام على تهجى حتى الكلمات البسيطة بصورة صحيحة مثل: Haemoglobin و Labour. وهذه المشكلة تصبح مركبة فى استراليا، حيث يوجد هناك موقف متناقض تجاه الكلمات التى تنتهى بـ our -؛ والهجاء المفضل كما يظهر فى the Macquorie dictionary هو our -، ولكن أحد

الأحزاب السياسية الكبرى هو حزب العمال Labor Party، ومنذ أكثر من مائة سنة أطلق على أحد الموانئ في جنوب استراليا Victor Harbor !

وإن الأخطاء سوف تؤثر على الصلة، من حيث أننا سوف نتلقى إجابات خاطئة؛ وسوف تؤثر كذلك على الاستدعاء، إذ أننا سوف تضيع علينا مواد كان يجب أن نجدها. ولذلك فينبغي أن نتأكد أن النظام الذى نستخدمه لا ينطوى فى بنيته على اتجاه لزيادة الخطأ البشرى. وكما أن الأخطاء فى خط تليفونى تقدم ضوضاء مسموعة وهذا يتداخل مع استقبالنا للرسالة؛ فإن الأخطاء البشرية فى نظام استرجاع المعلومات تقدم نوع الضوضاء الخاصة بها. وكلما كانت الأخطاء قليلة، كلما قلت الضوضاء من هذا المصدر؛ وهناك، مع ذلك، مصادر أخرى للضوضاء يبنى معالجتها، ومعظمها غير قابل للتحسين مثل هذا النوع.

سهولة الاستفادة

أياماً كان النظام الذى نختاره، فإن ثمة شخصين يبنى أن يجدها سهل الاستعمال: الشخص المسئول عن المدخلات، أى المكشف، والشخص الذى يحاول الحصول على المخرجات، أى المستفيد. فما القدر من المهارة الذى يحتاجه المكشف كى يكون قادراً على استخدام النظام فى مرحلة المدخلات؟ هل يساعد النظام فى التغلب على عجزه عن فهم الموضوعات التى يعالجها؟ فنحن لانستطيع جميعاً أن نكون أصحاب معرفة غير محدودة! والمستفيدون يصعب عليهم غالباً أن يُعبروا عن حاجاتهم على وجه الدقة؛ هل يساعدهم النظام رغم هذا على تكوين بحث مرض؟ هل الشكل المادى للمخرجات مقبول؟ فإن نظاماً يقدم للمستفيد مجموعة من الوثائق، أو على الأقل مستخلصات، أقرب إلى أن يكون مألوفاً وشعبياً عن نظام يعطى فقط سلسلة من الأرقام.

الخصوصية والشمول

هناك عوامل أخرى كثيرة تؤثر على أداء نظام استرجاع المعلومات عموماً وعلى إمكانياته من حيث الاستدعاء والصلة. ويمكن أن نتناول أولاً:

الخصوصية specificity: أى الذى الذى يسمح لنا به النظام من حيث دقة تخصيص موضوع الوثيقة التى نبحثها فكلما كانت قدرتنا على التخصيص أكبر، كلما كنا أقدر على تحقيق درجة أعظم من الصلة، والعكس، فإن النظام الذى يسمح لنا بدرجة محدودة من التحقيق أو الدقة فى التخصيص، سوف يمكننا من تحقيق درجة عالية من الاستدعاء، ولكن من درجة منخفضة نسبياً من الصلة. وفى المثال السابق، إذا لم يمكننا نظامنا من تخصيص القطط السيامية، فسوف يكون علينا أن نبحث فى كل الوحدات التى تتناول القطط قبل أن نتمكن من معرفة هل لدينا شىء عن هذا الصنف. بل أكثر من ذلك، إذا اتضح لنا أن الوحدة الثانية تتصل بالقطط السيامية، فلن يكون ثمة ضمان بأن هذه هى الوحدة الوحيدة، أو أن هناك وحدات أخرى معها فإذا كان النظام يفتقر إلى الخصوصية، فمعنى ذلك أننا قد عدنا إلى نوع من البحث المتتابع أو المتعاقب الذى نحتاج إليه إذا لم تكن مجموعتنا منظمة على الإطلاق باستثناء أننا قد ضغطنا طبعاً حجم المادة التى علينا أن نبحثها عن طريق تخصيص محتواها بصورة جزئية. فإذا أردنا أن نمارس أقصى درجة من السيطرة على بحثنا، فيجب أن يسمح لنا النظام بأن نخصص موضوعاتنا بدقة؛ وفى الحقيقة ينبغي أن يكون تخصيصنا فى كل حالة متطابقاً مع موضوع الوثيقة. فإذا أردنا أن نزيد من الاستدعاء، فإننا نستطيع دائماً أن نتجاهل جزءاً من التخصيص، ولكننا لانستطيع أن نزيد الصلة عن طريق الإضافة إليه (التخصيص) فى مرحلة البحث. ومن المهم جداً أن نفرق تفرقة واضحة بين المدخلات إلى نظامنا (أى ما نضيفه من تخصيصات الوثائق) وبين المخرجات (أى نتائج الفحوص التى أجريناها وسط هذه التخصيصات). ونحن لانستطيع أن نضيف إلى المدخلات فى مرحلة المخرجات؛ ولذلك فإن أى شىء يحذف فى مرحلة المدخلات سوف يبقى خارج النظام، وحينئذ يكون لزاماً علينا أن نستبدله بالبحث المتتابع لمخرجات ضخمة بدون داع (ومع ذلك فقد نكون قادرين على استخدام طريقة مقننة لكى تساعدنا فى تكوين إستراتيجيتنا للبحث، حتى ولو لم

تكن جزءاً من النظام المستخدم لتخصيص المدخلات، في حين أن استخدام الحاسبات قد يمكننا من الدوران حول المشكلة من خلال استخدام النصوص الكاملة كمدخلات، ومن ثم نقل كل عملياتنا التشفيرية، إلى مرحلة المخرجات. هذه الأفكار تناقش بالتفصيل بعد).

والخصوصية وظيفة للنظام، ولكن ثمة عاملاً آخر هو الشمول-exhaustivity وهو نتيجة قرار إدارى. وهذا هو المدى الذى نذهب إليه فى تحليل أية وثيقة لكى نحدد على وجه الدقة ما المحتوى الموضوعى الذى يجب أن نخصه. ويمكننا أن نميز بين الفكرة التى تشتمل عليها الوثيقة ككل، وبين الأفكار الفرعية التى قد تشتمل عليها؛ مثال ذلك: قد يهتم وصف لتجربة علمية بالغرض والنتائج بصورة كلية، ولكنه قد يحوى أيضاً وصفاً للأجهزة المستعملة. قد نقنع أنفسنا فى مكتبة شاملة كبيرة بأن نخصص الأفكار إجمالاً، فنعطى متوسطاً يعادل واحداً إلى واحد ونصف تخصيص لكل وثيقة، بينما فى مكتبة متخصصة، وبسبب الرغبة فى إستغلال المجموعات إلى أقصى فائدة لها، قد نرغب فى تكشيف الأفكار الفرعية أيضاً، بحيث نعطي عشرات من التخصيصات لكل وثيقة. ويعرف هذا بالتكشيف العميق، على عكس التلخيص فى الطريقة الأولى. قد يتطلب هذا أن نقول إن كتاباً ما هو عن درايدن Dryden، واىكرلى Wycherley، كونجريف Congreve، فانبروج Vanbrugh، وفاركهار Farquhar، بينما نحن فى التلخيص قد نكتفى بالقول بأنه عن مسرحية البعث. ومن الواضح أنه قد يكون من الصعب إن لم يكن من المستحيل أن نكشف كل الأفكار الفرعية فى وثيقة ما فى فهرس المكتبة أو فى بليوجرافية (يشتمل كشاف هذا الكتاب على ألف مدخل تقريباً)، وعلى هذا فإن التكشيف العميق ينفذ عادة فى المكتبات التى يمكن فيها توقع حاجات القراء بدرجة كافية من الوضوح، ويطبق التكشيف العميق غالباً على التقارير الفنية وما يشابهها من وثائق، وهى التى تكون قصيرة نسبياً ولذلك يمكن السيطرة عليها. وقد أنجز شىء من العمل عن إمكانية إستخدام قوائم محتويات الكتب وكشافتها لكى نزيد بشكل كبير جدا

من شمول التكشيف الممكن فى النظم المباشرة باستخدام الحد الأدنى من الجهد^(١٦). ولما كان جزء كبير جدا من الكتب المنتجة حاليا تجمع حروفها بواسطة الحاسب فإن النسخ المقروءة آليا فإن النسخ المقروءة آليا يمكن الحصول عليها بسهولة.

ويتضمنه التكشيف العميق أن يقوم المشكف بممارسة إصدار الأحكام على الأفكار والأفكار الفرعية وهل هى جديرة بالملاحظة والتسجيل أم لا. وعند الاختيار بين التلخيص والتكشيف العميق، فإن القرار هو قرارنا نحن وليس وظيفة للنظام. ومع ذلك فقد أظهر البحث الجديد أن ثمة رابطة بين الشمول والخصوصية، من حيث أن قيمة زيادة التخصيص تكون محدودة ما لم يكن النظام المستخدم يدبر خصوصية كافية، بمعنى أن التكشيف لن يحسن من طريقة الوصول إلى محتويات وثيقة ما مالم تكن المداخل الإضافية مخصصة.

وإن لحظة من التفكير توضح لنا أنه فى حين أن الخصوصية هى وسيلة لزيادة الصلة على حساب الاستدعاء، فإن الشمول يعمل فى الإتجاه المضاد، أى يزيد الاستدعاء ولكن على حساب الصلة. ويمكن أن نستخدم وسيلة لإبطال هذا الأثر إلى حد ما هى الوزن. وفيه نحاول أن نوضح أهمية أى تخصيص معين عن طريق إعطائه وزناً (ثقلاً) حسب مقياس سبق تحديده. مثال ذلك: إذا كان لدينا كتاب عن الحيوانات الأليفة المدللة يعالج الكلاب بصورة موسعة يمكننا أن نعطى الحيوانات الأليفة ثقلاً من $\frac{1}{10}$ ، والكلاب $\frac{1}{8}$ أو أقل. فإذا كان يقدم بعض المعلومات عن الكلاب، ولكنها معلومات قليلة، يمكن أن نعطى الكلاب وزناً أو ثقلاً منخفضاً من $\frac{2}{10}$. وحينئذ فإن القارئ الذى يريد صلة عالية يعرف أن بوسعه أن يغفل هذه الوحدة (المادة) بالذات، على الأقل فى الوقت الراهن، بينما القارئ الذى يريد استدعاء عالياً لن يجد صعوبة فى التعرف على هذه الوحدة. ويمكن وزن مصطلحات البحث بالطريقة نفسها، ويشتمل نظام سمارت SMART على طرق لاستخدام التغذية الراجعة من أحكام المستفيد

بالنسبة للصلة لكي يغير استراتيجية البحث من خلال تغيير الوزن المعطى لمختلف مصطلحات البحث.

الوقت

يستغرق التكشيف وقتاً؛ وكذلك البحث. وعن طريق زيادة جهدنا في مرحلة التكشيف - أي المدخلات - فقد نصبح قادرين على تقليل مقدار الوقت الذي علينا أن نقضيه في مرحلة المخرجات في البحث. ومن ناحية أخرى، ففي أي وضع مكتبي، هناك جزء من الوثائق المكشوفة (قد يكون كبيراً) لن يبحث أبداً، والجهد الذي يبذل في تكشيفه ضائع؛ فإذا ركزنا جهدنا في مرحلة المخرجات عن طريق ضغط تكشيفنا إلى الحد الأدنى (مثال ذلك: عن طريق إستخدام المستخلصات أو حتى النصوص الكاملة لمرحلة المدخلات بدلا من تخصيصات الموضوعات) ثم نجرى بحثاً معقداً لكي نستخرج الوحدات المتصلة، لأمكن أن نجادل بأننا قد وفرنا قدراً كبيراً من العمل غير الضروري. وكما ذكرنا آنفاً، فإن القراء لا يستطيعون أن يحددوا ما يريدون على وجه الدقة، ولذا فإن أي بحث سيكون حواراً بين المستفيد والنظام؛ وإن نتائج البحث الأول سوف تستخدم لتعديل أو تحسين السؤال بحيث يمكن إجراء البحوث التالية إلى أن يتم الوصول إلى وقت يمكن أن يعد نقطة نهاية مرضية. وعن طريق تركيز جهودنا على البحث وليس على التكشيف فنحن لانشوش على هذا الحوار بأية حال، ولكننا نستطيع بسهولة أن نستفيد من هذه التغذية الاسترجاعية في تخطيط إستراتيجيات البحث في المستقبل. وفي نظام تركز فيه الجهود على المدخلات، بصورة راجعة؛ فإذا أردنا أن نضع في اعتبارنا الخبرة المتحصلة من البحث فقد يعنى ذلك أن علينا أن نعيد تكشيف بعض الوحدات - أي نزيد من الجهد في المدخلات إلى مدى أبعد.

وفي الوقت الحاضر تنطوي كل الأنظمة تقريباً على مقادير كبيرة من الجهد في مرحلة المدخلات بدلا من تحويل ذلك إلى مرحلة المخرجات. وإن إستراتيجيات البحث المعقدة التي لازال يتطلبها الإجراء الأخير حتى الآن، إلى

جانب المقادير الكبيرة من المعلومات التي تطلب، هذه وتلك تجعل منه إجراء غير عملي في الأوضاع المكتبية المعتادة، رغم أن هذا في سبيله إلى التغيير، إذ أن حاسبات الكترونية أقوى وأقدر تصبح الآن في حيز الاستعمال وإن بعض خدمات بام التي أشرنا إليها آنفاً تعتمد على المستخلصات أو حتى على العناوين وحدها، وإن نتائج هذا بالنسبة لبناء إستراتيجية البحث أو بيانات القراء سوف تناقش في وقت لاحق بمزيد من التفصيل. ولازال بحث النصوص الكاملة مكلفاً إلى حد يصعب تبنيه بصورة واسعة لسنوات قادمة، وحتى في كل الحالات فلازال ينزع إلى تقديم المشكلات فيما يتعلق بالصلة، طالما أن شمول الكشف سوف يكون بطبيعة الحال ١٠٠٪.

البحث التكرارى والموجه

إن فكرة الحوار بين المستفيد والنظام جديدة بالمزيد من الدراسة. فكما أشرنا آنفاً، يجد المستفيدون غالباً صعوبة في التعبير عن حاجاتهم بدقة. وفي مكتبة تقليدية، يقوم المستفيد بالبحوث بنفسه أو يقوم بها المكتبي نيابة عنه. فإذا قام المستفيد نفسه بالبحث، فسوف يتعدل البحث كلما تقدم؛ فكل وثيقة يجدها المستفيد تؤثر على قراره فيما يتعلق بالمعلومات التي يطلبها. وفي كثير من الحالات تتضح الصورة بينما البحث مستمر، وهذا التوضيح يؤدي إلى وضع ينتهي فيه المستفيد إلى هدف يختلف تماماً عن الهدف الذى بدأ به. ويمكن أن نصف هذا البحث الذى يتعدل وفقاً لمجرى الأحداث بصفة مستمرة في ضوء المعرفة الحاصلة، يمكن أن نصفه بالبحث الموجه heuristic. أما إذا قام المكتبي بالبحث، فلن يكون هذا التعديل المستمر ممكناً، إذ أن التغييرات التي تطرأ على معرفة المكتبي لا تؤثر على المستفيد. ولهذا السبب فإن المعتاد بالنسبة للمكتبي أن يقوم ببحث أولى ويقدم النتائج للمستفيد، وحيث يمكن أن تتعدل استراتيجية البحث في ضوء مقدار الوثائق المتصلة من بين تلك التي تنتج عن هذا البحث. وحينذاك يمكن أن يجرى بحث ثان، وتكرر العملية إلى أن يصل المستفيد إلى ما يريد. هذا النوع من البحث الذى يعدل على فترات وليس بصفة مستمرة، يوصف بأنه تكرارى iterative. وكل من النوعين من البحث: التكرارى

والموجه يتطلب تفاعلاً بين المستفيد والنتائج، ولكن البحث الموجه يضغط وقت الإنتظار بين تسلم نتيجة عملية البحث وبين إستخدامها فى تعديل طريق البحث.

وإن الكثير من نظم استرجاع المعلومات لا يسمح بالبحث الموجه، مثل معظم الأنظمة التى تناولها الجزء الرابع من هذا الكتاب، بينما فهرس المكتبة البطاقى التقليدى يسمح به. ولاينبغى أن نبالغ فى تقدير هذا ولكن من الواضح أنه نقطة جديرة بالاعتبار عند محاولتنا لتقدير القيمة النسبية للأنظمة المختلفة^(١٧).

الاستطلاع

كان افتراضنا حتى الآن هو أن الغرض من النظام الذى نستخدمه هو أن يجعل من الممكن إيجاد المعلومات عند طلبها - أى أن القراء سوف يأتون إليه بهدف محدد فى أذهانهم، وحتى إذا لم يكن ذلك كذلك، فإن البدء فى البحث سوف يوضح هذا الهدف. ومع ذلك، فليست هذه أبداً هى الحالة عادة؛ ستكون هناك مناسبات يأتى القراء فيها إلى المجموعات دون حاجة معينة فى الذهن، ولكنهم، بدلا من ذلك، يرغبون فى التقاط وحدات بطريقة عرضية. ولكى يكون النظام قادراً على مساعدتهم فى مثل هذه الظروف فلا بد أن يسمح بالاستطلاع-Brows-ing؛ أى أن يكون باستطاعة القارئ أن يتتبع مجموعة عرضية من الأفكار كما يستطيع إجراء البحث الموجه. وكما أشرنا عند مناقشة بام، فإن الوحدة التى يثبت غالباً فيما بعد أنها الأكثر أهمية لاتوافق فى البداية مع نماذج إهتمامنا؛ والكثير من أهم الاكتشافات العلمية قد جاءت نتيجة السرنديبية - «موهبة أن تصبح سعيداً وأن تقوم باكتشافات غير متوقعة بالمصادفة»^(١٨)- وإن النظام الذى يستبعد هذه الإمكانية قد يثبت أنه أنجح مما يجب فى التوافق مع الحاجات التى يعبر عنها القراء!

التكلفة

إن عوامل التكلفة من أكثر العوامل تأثيراً على نظم إسترجاع المعلومات

وينبغي أن نوازن بين تكاليف أن ننظم مكتباتنا بحيث نتمكن من إيجاد المعلومات عندما تطلب، وبين ألا نجدها على الإطلاق، أو نجدها متأخرة بعد فوات أوان الاستفادة بها. وفي المكتبات التي تخدم الشركات الصناعية مثلاً، قد تكون تكاليف عدم إيجاد المعلومات عالية جداً؛ وهذا هو السبب مثلاً، في أن «رجال الأعمال العمليين» يضيفون إلى أعبائهم بأن يدفعوا تكاليف خدمات مكتبية غالية. (يقصد بالمصطلح «الخدمات المكتبية» هنا أن يشتمل على ما يدل عليه المصطلح «خدمات المعلومات» الذي يستخدم في مجال الصناعة). ومن جهة أخرى، فإن المكتبات العامة نزعَت في الماضي إلى اعتبار أن استغلال المعلومات التي تحويها مجموعاتنا أقل أهمية بكثير جداً من مجرد تدبير هذه المجموعات، وذلك لأن التكاليف على مستوى المجتمع إذا أخفق أحد أفرادها في الوصول إلى المعلومات التي يريدونها أقل بكثير جداً من تكلفة تنظيم المواد بصورة مناسبة. ومع ذلك فقد أصبح مفهوماً الآن أن تكاليف المعلومات الفاقدة على مستوى المجتمع مرتفعة جداً في الحقيقة بلغة المنافسة الدولية، ويكسر مزيد من الجهد لتوفير خدمات أكثر كفاية. ولا زال أمامنا أن نعرف الكثير عن تكاليف فاعلية مختلف طرق تنظيم المعلومات رغم أننا قد بدأنا نعرف شيئاً عن فعاليتها النسبية كأنظمة. وعلى الرغم من جهلنا النسبي فيجب ألا نغفل عوامل التكلفة كلية، ولكن هذه العوامل لا يمكن أن تدرس بالتفصيل عادة إلا وسط مجموعة من الظروف، ولذلك فسوف نتناولها في هذا النص في ألفاظ عامة.

تميل الإتجاهات الحديثة في تقويم أهمية التكلفة إلى أن تكون في صالح فكرة فاعلية التكلفة. وقد كان الهدف من معظم الوسائل المعقدة التي طورت في السنوات الأخيرة هو تحسين الصلة: تقليل عدد الوثائق غير المطلوبة التي يكشف البحث عنها، وهكذا يقل أيضاً الوقت الذي يستغرقه البحث خلال النتائج واختيار الوثائق التي تكون مفيدة لنا. ومع ذلك، فإذا كان إستخدام نظام معقد للتكشيف يتكلف أكثر من البحث في مخرجات نظام غير معقد، فليس ثمة داع لاستخدام نظام متقدم. كذلك ينبغي أن نحمل في أذهاننا أن مستوى الصلة

الذى يمكن قبوله فى نظام صغير قد لا يمكن قبوله أبدا فى نظام كبير على مستوى القطر يستخدم الطرق الآلية. وإذا كشف البحث عن عشر وثائق، أربع منها مفيدة، فإن هذا ليس شيئاً سيئاً جداً؛ ولكن إذا كان عندنا المستوى نفسه فى مجموعة تكبير هذه مائة مرة، فإننا نجفل رعباً عند فكرة إستبعاد ستمائة وثيقة من ألف. وحتى الآن لم يتم إلا قدر محدود من البحث حول هذا الجانب من إسترجاع المعلومات، ولكن من الواضح أنه مجال لمزيد من الاستكشاف فى المستقبل، وخاصة مع تطور الأنظمة الآلية.

مشكلات الترتيب الطولى

المعرفة متعددة الأبعاد: أى أن الموضوعات مرتبطة ببعضها بطرق متعددة. وقد افترضنا، فى المثال الذى أعطيناه من قبل، أن القلط السيامية قد اعتبرت حيوانات أليفة مدللة، ولكن من الواضح أنه يمكن النظر إليها بعدة طرق أخرى - كفرع من الرتبة الحيوانية: السنوريات (الهررة - القلط - والأسود والنمور، إلخ)، أو أن أصلها فى جزء معين من العالم، ولنكتف بهاتين الناحيتين فقط. ومع ذلك فإذا حاولنا أن نرتب الوحدات فى مكتبتنا أو فى فهرسنا، فسوف نجد أننا محدودون بتسلسل طولى خطى linear، ذى بعد واحد، تماماً كما نفعل حينما نقرأ كتاباً. فنحن لانستطيع أن نعرض العلاقات المتعددة، ولذلك فيجب أن نجد طريقة أخرى لعرضها. فإذا كان لدينا كتاب بدون قائمة محتويات أو كشف، فإن الطريقة الوحيدة التى يمكن بواسطتها أن نجد جزئية ما فيه هى أن نقرأ فيه (حتى نصل إليها). ومع ذلك يمكننا التغلب على هذه المشكلة عن طريق توفير طرق متعددة للوصول خلال قائمة المحتويات والكشاف، وهذا يسمح لنا بالذهاب مباشرة إلى المعلومات التى نطلبها؛ ولكن نص الكتاب لازال يعرض معلوماته فى بعد واحد. والتسلسل الذى فى الكتاب قد اختاره لنا مؤلفه ولانستطيع تغييره، مع أننا يمكن أن نقلل من أثره إلى حد بعيد عن طريق تسجيل علامة كافية فى صورة كشافات وإرشاد.

ونحن نواجه المشكلة نفسها بالضبط فى تنظيم المعلومات فى مكتباتنا. فنحن نستطيع توفير تسلسل نأمل أن يكون مفيداً لقرائنا، كما يفعل المؤلف تماماً، ولكن يجب علينا أن ندرك الحاجة إلى الوفاء بالطرق الأخرى للوصول ويجب أن نعى أيضاً أنه بدون هذه الطرق الثانوية للوصول فإننا سوف نجد المعلومات من خلال طريقة واحدة فقط، اللهم إلا إذا كنا مستعدين للتحويل إلى البحث المتتابع. وثمة مثال بسيط سوف يوضح هذا بالنسبة لأداة مألوفة، هى دليل التليفونات. فهذه الأدلة ترتب ألفبائياً حسب أسماء عائلات المشتركين. فإذا كنا نعرف اسم المشترك، فسوف نجد رقم تليفونه دون عناء، ولكننا لانستطيع أن نؤدى العملية بالعكس؛ أى أننا لانستطيع أن نجد إسم مشترك نعرف رقمه، اللهم إلا إذا كنا على إستعداد للبحث فى الدليل حتى نجده. وللتغلب على هذا، يمكننا أن نعد تسلسلا آخر، يرتب هذه المرة حسب الرقم؛ ولكننا لايمكن أيضاً أن نجد رقم صديق إذا كنا نعرف فقط اسمه الأول وعنوانه.

والمشكلة بطبيعة الحال اقتصادية إلى حد كبير. فنحن لانستطيع أن نقدم تسلسلات متعددة للكتب والوحدات الأخرى فى مكتباتنا لأن محاولة ترتيب نسخة من كتاب ما فى كل نقطة فى المكتبة يكون لهذا الكتاب صلة بالوحدات الأخرى فيها، هذه المحاولة تكلفنا أكثر من اللازم. ولانستطيع كذلك أن نعد تسلسلات متعددة فى الأدوات البيلوجرافية التى تعد للطبع والتوزيع. ربما استطعنا أن نعد مداخل عديدة فى سجلاتنا التى تحفظ فى داخل المكتبة، ولكن حتى هذا سوف يثبت أنه مكلف جداً إذا أردنا أن نحقق الاطراد والشمول. ومع ذلك فكما نستطيع التغلب على المشكلة فى الكتاب من خلال تدبير طرق الوصول المتعددة عن طريق التسلسلات الثانوية التى تقودنا إلى النقاط المطلوبة فى تسلسلنا الرئيسى، نستطيع أن نفعل الشئ نفسه فى نظام استرجاع المعلومات. والنظم المختلفة سوف تسمح لنا بدرجات متفاوتة من طرق الوصول المتعددة؛ وكلما كان النظام أكثر مرونة فى هذا الخصوص كلما كان أكثر قيمة. وربما كانت هذه أهم ميزة فى النظم المبنية على الحاسب. وحالما

تخزن المعلومات فى الآلة، فسوف يكون بإمكاننا أن نتداولها حقيقة بأى عدد من الطرق نرغب فىه؛ لدينا كل المرونة التى نرغبها دائماً ولكن لا يمكن الحصول عليها مع الأنظمة اليدوية.

السند الأدبى:

أياً ما كان النظام الذى نستخدمه، فإن المعلومات التى يشتمل عليها ينبغى أن تكون وظيفة للمدخلات، أى أن نظامنا ينبغى أن تدخل فى حسابها العلاقات بين الموضوعات والتى تعرضها الوحدات التى نقوم بتكثيفها. وبالإضافة إلى ذلك، ينبغى أن نبني فيها العلاقات التى ندرکها بطريقة استنتاجية بين الموضوعات، من خلال دراسة جوهرية للمعرفة نفسها، ولكن إذا اقتصرنا على دراسة المعرفة وحدها دون أن نأخذ فى الحسبان المعرفة كما هى فى شكلها المدون، فسوف نجد أنفسنا غير قادرين على تخصيص الموضوعات بدقة. فنحن، بمعنى آخر، نهتم بتنظيم المعرفة فى المكتبات، وليس المعرفة كمعرفة ويستخدم المصطلح السند الأدبى Literaty warrant هنا بمعنى أن على نظامنا أن يعتمد على المادة التى نضعها فيه وليس على اعتبارات نظرية خالصة. (لما كان هذا المصطلح يستخدم فى هذا السياق على نطاق واسع، فإنه قد بقى، حتى رغم أن الأنظمة المبنية على الحاسب قد تحتفظ بمعلومات لا يمكن أن تتاح فى أى شكل آخر، والأنظمة التقليدية قد تحتفظ بمعلومات ترتبط بالمواد السمعية البصرية والتى لا تتناسب مع التعريف المعتاد «للأدبى»).

وثمة جانب آخر لهذا السؤال الخاص. فمادام الغرض الكلى من النظام هو المخرجات، فإن ذلك معناه أن مخرجات النظام هى التى تهتم. ولكننا لا يمكن أن نعرف سلفاً ما المخرجات التى سوف تطلب، أو على الأقل لا يمكن أن نعرف ذلك بدرجة ما من الدقة، رغم أننا يمكن أن نخمن تخميناً ذكياً على أساس من الخبرة السابقة. وهكذا، فمع أنه من المرغوب فيه أن نبني نظامنا بحيث يتوافق مع المخرجات المطلوبة. فلسنا قادرين على أن نفعل ذلك طالما أننا لانعرف

ماذا ستكون عليه المخرجات المطلوبة. فنحن مضطرون إلى أن نجعل المدخلات أساساً لنا في بناء نظامنا، مضافاً إليها ما تقترحه الدراسات التي نجريها عن المعرفة خارج النظام. ولو قصرنا أنفسنا على دراسات المعرفة خارج النظام، وذلك بإغفال المدخلات، فإننا بذلك نقضى نظامنا خطوة عن المخرجات المطلوبة. وثمة حصيلة من المعرفة في أي مجال موضوعي، ولكن أية وثيقة نقوم بتكسيهها قد تعدل هذا؛ والسند الأدبي يعني أن يكون النظام قادراً على استيعاب هذا النوع من التغيير.

وربما كان ثمة خطورة في أن نقف موقفاً سلبياً بالنسبة للسند الأدبي: نستبعد من نظامنا احتمال استيعاب الموضوعات التي لم تكن قد ظهرت بعد في مجموعتنا. وترتبط هذه الخطورة عادة بنوع النظام القديم الحاصر الذي سوف نتناوله بعد، ولكن ظهرت أمثلة أحدث تبين المشكلات التي تنشأ إذا جعلنا نظامنا استاتياً ثابتاً بصورة مقصودة. فإذا أردنا أن نحافظ على مستوى من الخصوصية كالذي وصفناه، فإن المرونة بمعنى قابلية استيعاب المفاهيم الجديدة التي تكشف عنها المجموعات تكون أمراً حيوياً.

وقد استعمل مصطلح السند الأدبي وندهام هلم، ويقصد به فكرة من نوع آخر، مع أنها مشابهة أساساً. فقد اعتبر أنه لو كان لدينا وثيقة عنوانها مثلاً: الحرارة، والضوء والصوت، فإن ذلك يمثل موضوعاً يجب أن نوفر له في نظامنا.

ومع ذلك، فإن تلك الموضوعات ليست موضوعات حقيقية، بل هي تجميعات للموضوعات تنتج عن الصدفة البليوجرافية، لكونها جمعت داخل جلد واحدة. ومن الأفضل أن تعالج على أنها موضوعات قائمة بذاتها وتكشف هكذا أيضاً. وينبغي ألا يختلط هذا الوضع مع التفاعل الحقيقي بين الموضوعات؛ مثال ذلك: أثر الحرارة على الصوت (نوقش بتفصيل أوسع فيما بعد)؛ وهذا نوع آخر من الأوضاع يجب أن ندبر له. ويندر أن نجد المصطلح

الآن بمفهوم هلم، رغم أن أفكاره قد انعكست إلى حد بعيد في عمل مكتبة الكونجرس*، ثم طورت بالفعل إلى النظرية الحديثة التي لخصناها فعلا.

الرأس والوصف :

نحن نستعمل المصطلحات في نظامنا للتكشيف لتسمية موضوعات الوثائق في مجموعاتنا، ولكن من الواضح أن المستفيد الذي وجد الوصف الموضوعي الصحيح سوف يطلب، علاوة على ذلك بعض التفاصيل عن الوثائق التي يصدق عليها ذلك الوصف. ولذلك يمكننا أن نقسم المدخل في النظام إلى جزأين : الرأس والوصف Heading and Description.

فالرأس هو الوصف الموضوعي الذي يحدد المكان الذي سوف نجد فيه مدخلا في التسلسل. (يقتصر كتابنا هذا على الاعتبارات الخاصة بطرق الوصول إلى الموضوع؛ أما في الفهرس الكامل فإن الرؤوس سوف تتضمن أسماء المؤلفين والعناوين إلى جانب الموضوعات). وتتألف الرؤوس، في النظام الأبائى، من الكلمات، أما الترتيب المقنن فيستخدم الرمز - اللغة الكودية - بدلا من الرؤوس.

ونحن نحتاج إلى أن نميز بين نوعين من الرؤوس. فلدينا الرؤوس التي هي مصطلحات «مفضلة»، بمعنى أننا نستخدمها لكي تقودنا مباشرة إلى المعلومات - مدخل فهرس، أو كتاب على الرف. وهذه تكون لغة الكشاف - index vocabulary. ولدينا أيضا المصطلحات غير المفضلة، مثال ذلك : المترادفات التي قررنا ألا نستخدمها، إلا من حيث أنها تقودنا إلى المصطلحات في لغة الكشاف. هذه المصطلحات غير المفضلة، مع لغة الكشاف، تكونان لغة المدخل؛ أما الوصف فهو الجزء من المدخل الذي يعطينا معلومات عن وثيقة ما، ولذلك

* يقال دائما أن تصنيف مكتبة الكونجرس قد تأثر بأفكار ونظريات هلم عن السند الأدبي، بمعنى أن الفريق الذي عمل في إعداد التصنيف قد تأثر بنظرية تصنيف الكتب، وليس بنظرية تصنيف المعرفة. والحقيقة أن هذا الفريق قد بدأ عمله في العقد الأخير من ق ١٩، وأن هلم قد بدأ مقالاته في الموضوع في ١٩١١ - ١٩١٢، ولذلك فإن تصنيف مكتبة الكونجرس وإن كان قد قام على أسس مشابهة إلا أنه أسبق (المرجى).

فسوف يشتمل على كل العوامل التي تفيد في تحقيق الذاتية. وثمة مجموعات متنوعة من القواعد الخاصة بجمع أوصاف الوثائق، مثل القواعد التي يشتمل عليها التقنين الأنجلو - أمريكي، أو التقنين الدولي الموحد للوصف الببليوجرافي، ولكن لانتاج لأغراضنا هنا إلى أكثر من الإشارة إلى أنها موجودة. وإن وجود وصف للوثيقة يمكننا من إجراء فصل مفيد : فالمدخل الموضوعي يتألف من رأس من ألفاظ الكشف مع وصف للوثيقة، في حين يقودنا مدخل الكشف أو الإحالة التبادلية من رأس بدون وصف للوثيقة إلى مدخل ما. والرأس الذي نعد منه إحالة قد يكون رأساً يظهر فقط في ألفاظ المدخل، وفي هذه الحالة سوف نعد إحالة أنظر، وهي تنقلنا من رأس غير مستعمل إلى رأس مستعمل؛ أو قد يظهر في ألفاظ المدخل والكشاف معاً، وفي هذه الحالة ستكون الإحالة أنظر أيضاً، وهي تربط رأسين كل منهما مستعمل، وذلك حتى توضح وجود نوع من الصلة.

ويجب ملاحظة أن الأوصاف في بعض النظم قد تكون في شكل رقم (رقم الورود أو رقم الوثيقة) لا في صورة معلومات مفصلة عن المؤلف، والعنوان، وبيانات النشر وما إلى ذلك، وهي المعلومات التي نجدتها في الفهرس البطاقي للمكتبة مثلاً. وقد تكون الروابط بين الرؤوس المتصلة جزءاً لا يتجزأ من التسلسل الرئيسي للمداخل، كما في الفهرس القاموسى، أو جزءاً من تسلسل إضافى، كما في الفهرس المصنف؛ أو قد تكون الروابط مستقلة تماماً كما هو معتاد في النظم اللاحقة. وهذه النقاط سوف تتضح كلما تقدمنا في هذا الكتاب، أما في هذه المرحلة فمن المهم أن ندرك أن هذه الخواص، مثلها مثل غيرها في هذا الفصل، مشتركة في كل أنظمة استرجاع المعلومات وإن وجودها أو غيابها يؤثر تأثيراً كبيراً في السهولة التي يمكن أن نسترجع بها المعلومات.

المدخل اللفظ والمدخل الوحدة

إن الفقرات السابقة تعنى أننا نعد مداخل للوثيقة (التي تحقق ذاتيتها عن طريق

وصفها) تحت كل من الرؤوس المناسبة، ونصف هذه الأوصاف في المكان الصحيح في تسلسلنا الألفبائي أو المصنف، والنظام الذي يعمل بهذه الطريقة يطلق عليه نظام المدخل اللفظ term entry والنظام البطاقي الذي يستخدم البطاقات الموحدة هو من هذا النوع. ومع ذلك فمن الممكن أن نتبنى الطريقة العكسية، فنعد مدخلا واحداً لكل وحدة أو جزئية، مستخدمين في ذلك نوعاً من الشكل المادى يسمح بالوصول إلى المداخل من خلال الرؤوس اللازمة جميعاً، ومثل هذا النظام يعرف بالمدخل الوحدة item entry وهو يستخدم في الأنظمة المبنية على الحاسب وهو يعنى ضمناً وجود تسلسل رئيسى من المداخل من مدعوم بكشاف أو أكثر، وقد ناقشناه بالتفصيل فى الفصل ٥.

فصل الجهد الذهني عن الجهد الكتابي

ينقسم العمل فى أى نظام إلى جزء ذهنى (عقلى) وجزء كتابى. وإن تحديد تحت أى الرؤوس تدخل الوثيقة هو عمل ذهنى فى نظم التكشيف المعنية والتي ناقشناها فى فصول تالية، ولكن العمليات الآلية الفعلية لإدخال الوثيقة تحت هذه الرؤوس فى الملف ليست عمليات ذهنية. وكذلك فى عملية البحث، فنحن نحتاج إلى اتخاذ قرار ذهنى وهو أى الرؤوس تكشف الإجابات عن سؤال معين، ولكن العمل نفسه وهو عرض نتائج البحث لا يتضمن جهداً ذهنياً.

وعندما يصبح الأمر متعلقاً باستخدام الحاسبات الإلكترونية فإن الفصل يصبح هاماً. فالحاسبات تستطيع أداء العمليات الكتابية الروتينية جيداً : فهى أدق، وأسرع من نظائرها البشرية، على فرض أنها غذيت بالتعليمات الصحيحة. نحن لانعرف فى الوقت الحاضر مايكفى عن الطريقة التى يعمل بها العقل البشرى لكى يكون قادراً على إعطاء الحاسبات التعليمات الصحيحة لتمكينها من أداء العمليات الذهنية؛ ولذلك فيجب أن تبقى هذه العمليات تتم بواسطة الجهد البشرى. ومع ذلك، فقد تمت دراسات متأنية كتلك التى قام بها ا.ج. كوتس E J Coates فى مجال استخدام الحاسب فى إنتاج كبت BTI، وقد أوضحت هذه

الدراسة أن كثيراً من العمليات التي كان يعتقد في الماضي أنها ذهنية يمكن في الحقيقة ضغطها إلى مجموعة من القواعد (الحسابية) التي يمكن أن تؤلى. ومن الواضح أن نقل أكبر قدر من العمل الروتيني إلى الآلات مفيد، فعن طريق هذا العمل فقط يمكننا أن نحسن خدمتنا للقراء.

المراجع

- 1 Vickery, B. C., *Techniques of information retrieval*. London, Butterworths, 1970. Chapters 1 and 2.
- 2 Orwell, G., 1984, London, Secker & Warburg, 1949.
- 3 Kemp, A., *Current awareness services*, London, Bingley, 1979.
- 4 The phrase 'one skilled in the art' is commonly used in patent specifications to denote someone who has a sufficient working knowledge of the existing procedures to be able to utilize the invention being patented.
- 5 Cleverdon, C. W., Mills, J., and Keen, E. M., *Factors determining the performance of indexing systems*. Cranfield, Aslib-Cranfield Research Project, 1966, 2v in 3.
- 6 Lancaster, F. W., 'Evaluation and testing of information retrieval systems', in *Encyclopedia of library and information science*, 8, 1972, 234-59.
Lancaster, F. W., 'Pertinence and relevance', in *Encyclopedia of library and information science*, 22, 1977, 70-86.
Lancaster, F. W., 'Precision and recall', in *Encyclopedia of library and information science*, 23, 170-80, 1978.
Swanson, D. R., 'Subjective versus objective relevance in bibliographic retrieval systems', *Library quarterly*, 56 (4), 1986, 389-98.
For a recent detailed review of the question of relevance see Schamber, L., 'Relevance and information behavior', *Annual review of information science and technology*, 29, 1994, 3-48.
- 7 Cleverdon, C. W., *Aslib Cranfield Research Project: report on the testing and analysis of an investigation into the comparative efficiency of indexing systems*, Cranfield, College of Aeronautics, 1962.
- 8 Buckland, M. and Gey, F., 'The relationship between recall and precision', *Journal of the American society for information science*, 45, 1994, 12-19.
- 9 Zadeh, L. A., 'Fuzzy sets', *Information and control*, 8, 1965, 338-53.
Robertson, T. E., 'On the nature of fuzz: a diatribe', *Journal of the American Society for Information Science*, 29 (6), 1978, 304-7.
Cerny, B., 'A reply to Robertson's diatribe on the nature of fuzz', *Journal of the American Society for Information Science*, 30 (6), 1979, 357-8.
Bookstein, A., 'Probability and fuzzy set applications to information retrieval', *Annual review of information science and technology*, 20, 1985, 117-51.
- 10 I am indebted to E. M. Keen for drawing my attention to this point in ref. 5 above.

- 11 Fairthorne, R., 'Automatic retrieval of recorded information', *Computer journal*, 1958, 36-41.
- 12 Shaw, R., *private communication*, quoted by Cleverdon in *Journal of documentation*, **30** (2), June 1974, 174.
- 13 Blair, D. C., 'Searching biases in large interactive document retrieval systems', *Journal of the American Society for Information Science*, **17** (3), 1980, 271-7.
- 14 Lantz, B. E., 'The relationship between documents read and relevant references retrieved as effectiveness measures', *Journal of documentation*, **37** (3), 1981, 134-45.
- 15 Conrad, R. and Hille, B. A., 'Memory for long telephone numbers', *Post Office telecommunications journal*, **10**, 1957, 37-9.
- 16 Atherton, P., *Books are for use: final report of the Subject Access Project to the Council on Library Resources*, Syracuse, NY, Syracuse University, 1978.
Cochrane, P. A., *Redesign of catalogs and indexes for improved online subject access: selected papers of Pauline A. Cochrane*, Phoenix, AZ, Oryx Press, 1985.
- 17 Lancaster, F. W., 'Interaction between requesters and a large mechanized retrieval system', *Information storage and retrieval*, **4** (2), 1968, 239-52.
- 18 Hulme, E. Wyndham, *Principles of book classification*, London, Association of Assistant Librarians, 1950 (AAL Reprints No 1). Originally published in the *Library Association record*, 1911-1912. Included in *Theory of subject analysis* . . .

التكشيف المشتق ١ : الكشافات المطبوعة

رأينا فى الفصل ١ ، أنه يجب علينا أن نشفر موضوع الوثيقة حتى يتسنى لنا أن نضع الوثيقة نفسها أو تسجيلاتنا الخاصة بها فى خزانتنا. هذا يعنى أن نكون قادرين بطريقة ما على تخصيص الموضوع. كيف يمكن لنا أن نحدد موضوع وثيقة ما بحيث نكون قادرين على تخصيصه؟ لن يكون لدينا فى العادة وقت يسمح لنا بأن نقرأ كل الوثائق التى نضيفها إلى المجموعات، وعلى أية حال فقد لا نفهمها إذا قرأناها. سوف نستخدم مصادر مختصرة : صفحة المحتويات، التصدير أو المقدمة، أو وصف الناشر الذى يطبع على غلاف الكتاب؛ أو مستخلص إذا كنا نفحص مقالات فى مجلة أو تقريراً فنياً؛ أو الدعاوى الخاصة بمواصفة براءة اختراع. وهذه كلها سوف تعطى شيئاً من الدلالة على الموضوع وتقتراح خطوطاً معينة للتفكير إذا كنا نريد أن نتابع المسألة بصورة أبعد من ذلك، فى قاموس أو دائرة معارف مثلاً.

قد نقرر أنه بسبب التكلفة سوف نعلم فقط على المعلومات المعروضة فى الوثيقة، دون أن نحاول أن نضيف إليها من معرفتنا الخاصة أو من مصادر أخرى. وهذا هو التكشيف المشتق، أى : التكشيف الذى يشتق مباشرة من الوثيقة. ويمكننا أن نبدأ بدراسة بعض الطرق التى استخدم بها التكشيف المشتق لإنتاج كشافات مطبوعة، وبخاصة فى النظم المبنية على الحاسب. وتوجد هذه بكثرة فى الوقت الحاضر فى النظم على الخط المباشر، ولكن الأسس تبقى هى نفسها.

لقد رأينا من الممكن أن نميز بين الجهد الذهني والجهد الكتابي المتضمن في نظام استرجاع المعلومات، والحاسبات تمكنا من تنفيذ العمليات الكتابية بسرعة عالية. والتكشيف المشتق يقلل الجهد الذهني إلى الحد الأدنى، وهو لذلك يتناسب جيدا مع عمليات الحاسب، التي يمكن أن تمكنا من أن نحصل على مخرجات متنوعة من مدخلات واحدة. وقد نجد أننا قادرون على أن نتج بعض أشكال المخرجات التي يمكن إنتاجها يدويا ولكن بسبب عوامل الوقت والتكلفة لم يحاول أحد أن ينتجها^(١).

التكشيف المبني على العنوان

هناك بطبيعة الحال جزء من الوثيقة يحاول المؤلفون أنفسهم أن يعرفوا فيه الموضوع : العنوان. وفي حالات كثيرة يعطى العنوان دلالة واضحة عما تدور حوله الوثيقة، رغم أننا نجد حالات يتركنا العنوان فيها في شك، وحالات أخرى يصاغ فيها العنوان بحيث يجذب الانتباه أكثر مما يُعلمنا عن الموضوع. في الفئة الأولى يجب أن نضع :

The development of national library and information services;
Early Victorian New Zealand

وفي الفئة الثانية نضع :

The design of steel structures

(هل هي مبان؟ جسور؟ أم هي جميعا؟)، أو :

الترجمة الذاتية لفرد آستير! (Fred Astaire steps in time)

ونضع في الثالثة : Men in dark times (مجموعة من سير رجال ماتوا في

القرن العشرين، منهم برتولت بريخت Bertolt Brecht)

و "Waterfalls and tall building's"، الذي وجد أنه عرض لـ: كتاب جينيس

للأرقام القياسية! the Guinness book of records.

وينزع المؤلفون إلى التعميم في العناوين التي يختارونها، ورغم أنهم يحاولون عادة أن يجدوا عناوين متفردة بالنسبة لعملهم، فهذه ليست هي الحالة دائما.

هناك على سبيل المثال كتب متعددة تحمل عناوين غامضة مثل :

Materials and structures

والعنوان : *Malice in Wonderland* الذي استخدمه نيكولاس بلاك Nicholas Blake لقصة بوليسية قد استخدم أيضا لفيلم بطلته إليزابيث تايلور. وهكذا فإننا بصفة عامة نجد أن البحث عن عناوين مخصصة سوف يعطى استدعاء منخفضا، رغم أنه من المحتمل أن يعطى صلة عالية في الوقت نفسه، ولكنه سوف ينتج نقاطا زائفة *false drops* - عناوين تطابق التخصيص ولكنها ليست متصلة بآية حال .

وإذا نحن تناولنا العناوين المعطاة لأعمال جديدة، سوف نجد غالبا أن الأعمال عن الموضوع نفسه تحمل عناوين تحتوى على الكلمات المهمة نفسها - الكلمات المفاتيح - والتي يمكن أن تستخدم كأساس لاسترجاع المعلومات، مثال ذلك :

Manual of library classification

Library classification on the march

Introduction to library classification

A modern outline of library classification

Prolegmena to library classification.

وإن استخدام الكلمات المفاتيح لإنتاج أنواع مختلفة من الكشاف قد أصبح من الممارسات المستقرة، ولكنه تأكد في السنوات الأخيرة عن طريق استخدام الحاسبات لتداول المصطلحات .

تكشيف الكلمات الروابط Catchword indexing

استخدم تكشيف الكلمات الروابط لسنوات عدة في الأدوات الببليوجرافية، وبخاصة تلك التي ينتجها الناشر، حيث أنتج وسيلة رخيصة وفعالة إلى درجة معقولة للإتاحة الموضوعية للعناوين المسجلة (في الكشاف). وتعرض العناوين

بحيث تأتي في البداية الكلمات المهمة، وربما تعطى مدخلين أو ثلاثة لكل عنوان. ويختار المحرر الألفاظ (الكلمات) التي سوف تستخدم، وتُؤكِّدُ المداخل يدويا. وقد استخدم الأسلوب أيضا لإنتاج كشافات دوريات مثل : Nature. ومع تحسب هذه الأنواع من الكشافات، فلقد أسقط تكشيف الكلمات الرابطة في الواقع لصالح أشكال أخرى، ولكنه لا يزال موجودا في الأدوات المرجعية الأقدم.

تكشيف الكلمات المفاتيح في السياق (كويك KWIC)

طور ه.ب. لون H. P. Luhn من شركة ا ب م IBM^(٢) شكلا خاصا من أشكال تكشيف الكلمات الرابطة ينتجه الحاسب. وفيه تصبح كل كلمة دالة نقطة مدخل، ولكن بدلا عن ظهورها في الجانب الأيسر من الصفحة، تظهر الكلمة المفتاح في الوسط، مع بقية العنوان على كلا الجانبين. وهنا تظهر نقطة مهمة على الفور. في تكشيف الكلمات الرابطة، يختار المحرر المصطلحات المهمة، ولكن الحاسب لا يمكنه أن يتعرف أهمية لفظ ما؛ وبدلا من ذلك يتعين علينا أن نبني قائمة بالألفاظ التي ليست لها قيمة لأغراض التكشيف : قائمة توقف a stop list. وحينئذ يرمج الحاسب بأن يهمل أية مداخل تظهر تحت هذه المصطلحات. ومن المعتاد أن يكون هناك قائمة مختصرة إلى حد معقول بالمصطلحات التي يكون من الواضح أنه لا قيمة لها كمدخل كشاف - الأدوات : a, an, the ؛ والحروف : on, of, in ؛ حرفا العطف : and, or ؛ الضمائر : he, she, my ؛ وما إلى ذلك - ثم نضيف قائمة أكبر، مبنية على التجربة، بالألفاظ التي لا يتظر أن يكون لها أية قيمة.

وتشتمل دائرة معارف جروليير الجديدة متعددة الوسائط على قائمة توقف تضم ١٣٢ لفظا، وهي تشتمل على : بيليوغرافيا bibliography الأمر الذي لا يسعد المكتبيين! والمصطلحات التي اعتبرت غير مبحوث عنها هي ألفاظ مثل although, begun, can, different, etc. ويمكن أن تحرر القائمة من حين لآخر لكي يوضع في الحسبان «الجديد»، من المصطلحات والتغييرات في التغطية

الموضوعية للمجموعات ولازال هذا يتركنا مع ألفاظ ترد نادرا بدرجة تجعل من غير الممكن أن نضمناها قائمة توقف؛ وحينما ترد، فإنها سوف تؤدي إلى ظهور مداخل في الكشاف، ولكن هذه سوف تكون قليلة إلى درجة أنها لن تتعب المستفيد. ومن الأفضل أن تضمن مداخل قليلة ليست مفيدة على أن تخذف مداخل سوف تكون مفيدة.

وقد كانت أساليب كويك في وقت من الأوقات شائعة باعتبارها وسيلة رخيصة وسريعة لإنتاج الكشافات، ولكن مع الطرق الأكثر تعقيداً والمتاحة الآن، فقد حلت محل كويك. ومن الأماكن التي لايزال يوجد فيها مثل هذا الكشاف - كويك - المكتز، حيث يستخدم الكشاف الدائري rotated للكشف عن الألفاظ المحتجبة في المصطلحات المتعددة الألفاظ. وفيما يلي مقتطف من كشاف : the PAIS Subject headings list :

| | |
|----------------------|----------------------------------|
| | Industrial relations |
| Boycott | (industrial relations) |
| | Industrial relations consultants |
| Grievance procedures | (industrial relations) |
| | industrial safety |
| Social service, | industrial |
| Sociology, | industrial |
| Spies, | industrial |
| | industrial surveys |

ورغم أن الكلمة "Industrial" سوف تأتي في بداية عدد من هذه المداخل، أي : نقطة الإتاحة، فإنها سوف لا توجد في مداخل أخرى على الإطلاق بسهولة بطريقة أخرى غير هذا الكشاف. ومن الاستخدامات المهمة لأساليب كويك تلك التي قدمها ديالوج DIALOG⁽³⁾. وكثير من الملفات تحتوى الآن على ملايين عدة من المراجع، وحتى البحث الذي خطط له بعناية قد يسترجع عددا كبيرا من المراجع بدرجة غير مقبولة من هذا الملف الكبير. ولكي نسهل من عملية

فحص نتائج مثل هذا البحث، فمن الممكن عرض العناوين (أو النصوص في بعض الحالات) التي تحتوى على المصطلح المبحوث عنه في السياق. ويجب أن يكون هذا مفيدا في متابعة البحث بعد ذلك لتحسين أداء الصلة.

تكشيف الكلمات المفاتيح خارة السياق (كوك KWOC)

لما كانت الكلمة التي سيتم الترتيب على أساسها لا توجد في مكانها المناسب، لذا فإن تكشيف كوك يبدو غير مألوف، وهناك طريقة أخرى لتداول أو عرض العنوان هي أن نأتى بالكلمة المفتاح في بداية السطر، متبوعة بالعنوان الكامل. ولهذا ميزته، وهي أن يكون مظهره مألوفاً - الكلمة التي يتم الترتيب عليها تكون على اليسار - وكذلك فهو يأتي بالعنوان كله كما هو، ولكنه ليس ناجحاً مثل كويك الذى أتى معاً بالعناوين التي تحتوى على الأزواج نفسها من الكلمات. والفكرة موضحة جيداً بواسطة :

The British Library Document Supply Centre's Index of Conference proceedings

وقد بدأ هذا في سنة ١٩٦٥ بكشاف كوك ذى كلمة واحدة، ولكن كان عليه أن يطور نظاماً أكثر تعقيداً يستخدم أزواجا من الكلمات حيث نمت مجموعاته حتى وصلت إلى مايقارب ١٨٠٠٠ عنواناً كل سنة. مثال ذلك :

"A proposed new structure for food and agricultural policy"

«بنية جديدة مقترحة للسياسة الغذائية والزراعية»، يكشف تحت :

Food policy ؛ agricultural policy ؛ AAAS (المنظم)، ولكن ليس تحت

"Listening devices and citizen's rights : police powers and electronic surveillance".

تكشف تحت الأزواج من الكلمات :

Listening devices;

Citizens' rights;

Police Powers

Electronic Surveillance

و

ولا يبدو هذا أنه يحدد طريقة ما للبدء بالكلمة المفردة "Rights".

وباستخدام كويك وكووك، فإن كل عنوان سوف يؤدي إلى ظهور عدد من المداخل : بقدر ما هنالك من كلمات مهمة (دالة) في العنوان. ولهذا السبب، فإنهما يستخدمان عادة باعتبارهما كشافات، تقود إلى أوصاف الوثائق في ملف مستقل. وإذا حكمنا عليهما حسب معايير نظم استرجاع المعلومات بصفة عامة، فإن أداءهما ليس جيدا. من المؤكد أن الصلة سوف تكون عالية، من حيث أن العنوان الذي سوف نجده بالبحث في كلمة معينة - هذا العنوان سوف يكون مفيدا، ولكن قد يكون علنيا أن نبحث في عدد من المداخل تحت هذه الكلمة قبل أن نصل إلى العنوان الذي يشبه مانريده. ومع ذلك فالأرجح أن يكون الاستدعاء منخفضا. وقد رأينا أن المؤلفين يبحثون عن عناوين متفردة، وليس لدينا طريقة لتعرف المصطلحات المرتبطة، مثل المترادفات، اللهم إلا معرفتنا الشخصية، التي من الواضح أنها ستكون خارج النظام. وتعتمد الخصوصية على اختيار المؤلف للكلمات، في حين يعتمد الشمول مرة أخرى على درجة التفصيل في العناوين. ورغم هذا العيب المحتمل، فقد وجد أن مداخل كووك للعناوين كانت مألوفة لدى المستفيدين في :

the Bath University Comparative Cataloguing Study^(٤)، ولذلك فيجب أن يحمل على محمل الجد.

منذ سنة ١٩٦٠ فصاعدا، لقي مؤلفو البحوث العلمية والفنية التشجيع من الجمعيات المهنية والحكومة الأمريكية لكي يعطوا أعمالهم عناوين صريحة. ورغم أن الكشافات المطبوعة المبينة على العناوين ليست شائعة الآن، فإن العناوين مصدر مهم للمصطلحات في البحث في قواعد بيانات الحاسب في مجال العلم والتقانة، رغم أننا ربما نستثنى براءات الاختراع من هذا؛ فإن

عناوينها لابد أن تكون عبارات أو بيانات دقيقة عن المجال الموضوعي للمحتوى، ولكنها تصاغ بشكل غير مفيد لكي تتجنب مساعدة المنافسين. وفي العلوم الاجتماعية والإنسانيات هناك مشكلات المصطلحات التي تجعل عملية الكشف المبني على العنوان أقل فائدة.

الكشف الإسنادي (الاستشهادي)

إن الوثائق ذات القيمة تحتوي على بليوجرافيات؛ هذه هي الطريقة التي يبين بها المؤلفون الدعائم التي بنوا عليها. ويشير جارفي Garvey إلى أن قائمة المراجع هي جزء مفتاحي لأي ورقة علمية، حيث أنها تفيد في وضع البحث في سياقه الصحيح في تطور الأجماع العلمي⁽⁵⁾. ويقترح بحثه كذلك أن استخدام الإنتاج الفكري العلمي يحدث في مرحلتين منفصلتين تماماً من مشروع البحث. ومن المحتمل أن العلماء يبدأون أي عمل بحثي بالحصول على بعض المراجع، ولكن استعمالهم الرئيسي للإنتاج الفكري قد لا يحدث حتى يكتمل عملهم ويعد للنشر؛ في هذه المرحلة، يحاولون أن يبينوا صلة عملهم بالعمل التي تم من قبل، وربما عكست الأسانيد (الاستشهادات) هذا الأمر أكثر مما تدل على المصادر التي استخدمت فعلاً أثناء البحث. ويبدو أن هذا يعطى وزناً لأهمية الكشف الإسنادي Citation indexing.

هناك رابطة link بين وثيقة ما وكل عمل مسند في بليوجرافيتها، ويمكننا أن نعكس هذا فنقول إن هناك رابطة بين كل مادة مسندة في العمل والعمل الذي يسند إليها. ولما كانت الوثائق تسند في العادة إلى مواد متعددة، فإننا عن طريق بحث أعداد كبيرة من الوثائق الأصلية يمكننا أن نبني أعداداً كبيرة من هذه الروابط. فإذا رتبنا هذه حسب المواد المسندة، فسوف نأتى معاً بكل الوثائق التي اشتملت على مادة معينة في بليوجرافياتها. هذا هو المبدأ الأساسي في الكشف الإسنادي.

وكما هو الشأن مع الكشافات المبنية على العنوان، فإن استخدام الحاسب جعل ممكنا التداول العملى على مدى واسع لفكرة ما استخدمت بالفعل فى مجالات موضوعية معينة، وبدرجة ملحوظة فى الإنتاج الفكرى القانونى. ويتتج حاليا معهد المعلومات العلمية (ممع ISI)، الذى تأسس فى عام ١٩٦١، كشافات اسنادية متعددة، من أهمها :

Science citation index, 1961-

Social Science citation index, 1996-

Arts and humanities citation index, 1977-

وتغطى هذه الكشافات فيما بينها أكثر من ٥٠٠٠ دورية مفتاحية؛ وتبحث هذه، ثم تدخل كل الروابط البيولوجرافية الموجودة بها فى الحاسب. وتستخدم المعلومات التى تجمع على هذا النحو فى توفير الكشافات الاسنادية، وكشافات المصادر والكشافات المشتركة؛ كما تستخدم فى خدمة بام SDI، وآسكا ASCA (Automatic Subject Citation Alert)، و "كشاف موضوعى، هو the Permuterm index، الذى يدخل فى كل مادة تحت أزواج من الكلمات الدالة الموجودة فى العنوان^(٦).

ولكى نستخدم الكشاف الاسنادى، فمن الضرورى أن يكون لدينا المرجع لمقال متصل، ولكن يحدث كثيراً أن يبدأ البحث من أساس كهذا. فإذا لم يكن لدينا مقال «فورى»، فيمكننا أن نستخدم كشاف Permuterm لكى نحاول أن نحصل على نقطة بداية باستخدام أزواج من الكلمات المفاتيح من صياغة بحثنا؛ ويمكن أن يكون هذا مفيدا فى تحديد أماكن المعلومات عن موضوعات توصف بكثرة بواسطة زوج من الكلمات مثل "Holy Grail"، والذى قد يكون مستخدماً فى العنوان^(٧). وحينئذ يمكن أن يعطينا هذا مقالة أو أكثر لاستخدامها كنقطة بداية لبحث معين. ومن الكشاف الاسنادى يمكننا أن نجد تفاصيل موجزة لمقالات أخرى أحدث تكون أسندت إلى المقال الذى نعرفه فعلا. ويمكننا أن

نعود إلى كشف المصادر للحصول على تفاصيل كاملة، ومن ثم نحدد المقالات في الدوريات المناسبة. فإذا كانت غير متصلة فيمكننا استبعادها، ولكن إذا كانت متصلة فيمكننا استبعادها، ولكن إذا كانت متصلة فيمكننا أن نستخدم المقالات التي تسند إليها كأساس لمزيد من البحث في الكشف الإسنادى. ومن خلال عملية إعادة التدوير هذه recycling يمكننا أن نجتمع ببيوجرافية كبيرة من نقطة بداية مفردة. ومن الممكن طبعاً أن نفعل هذا يدوياً، ولكن هذا يتم بتكلفة عالية في الوقت والجهد. والبحث الذى يحتاج إلى أسابيع من العمل اليدوى يمكن عمله فى دقائق باستخدام الكشف الإسنادى.

ولما كانت كل مادة (وحدة) فى الدوريات المبحوثة يتم إدخالها، فإننا يمكن أن نتبع التغييرات والتصحيحات التى أجريت فى المقالات التى سبق نشرها. وتشتمل هذه غالباً على معلومات مهمة - مثال ذلك: محو نجاح مزعوم! - ولكن تتجاهلها عادة خدمات التكشيف والاستخلاص التقليدية. هذه الميزة ليست ملازمة (متأصلة) بطبيعة الحال فى التكشيف الإسنادى، ولكنها تتضح فى تكلفة كاع SCI ورفاقه مقارنة مع معظم الخدمات التقليدية.

روابط الاسناد

إن مقاليتين أسندت كل منهما إلى مقالة أسبق لابد أن يكون بينهما شىء مشترك؛ فإذا كان كل منهما يسند إلى مقاليتين سابقتين، فإن الرابطة تزيد. ويعرف هذا باسم الاقتران (التلازم) الببليوجرافى bibliographic coupling، وإذا كانت المقالتان تشتركان فى الاسناد إلى ستة مقالات، فإن هناك مايرر أن نفترض أنهما يغطيان الموضوع نفسه بدرجة عالية جداً. (لو أن مقاله تشترك فى الأسانيد ال ٥٠ مع مقالة أخرى فسوف تتحول هذه إلى أن تكون ترجمة!) وهذا يعكس حقيقة أن المؤلفين يسندون فى العادة إلى تلك الأعمال التى تكون الأساس الذى يبدأون منه كتاباتهم الخاصة. وقد أعطى الاقتران (التلازم)

الببليوجرافى نتائج طيبة فى الدراسات التى أجريت فى ممت MIT. وهناك مدخل ثبت أنه أعطى نتائج مفيدة هو الاسناد المشترك co - citation، الاسناد إلى مقالتين أو أكثر معافى أكثر من ورقة. مثال ذلك: حينما بدأت الدراسات الببليوجرافية تنتشر، كان من المؤلف أن نجد كتاب س. س. برادفورد Bradford عن التوثيق مسندا documentation، حيث أنه قد نشر فى هذا الكتاب أفكاره عن «قانون برادفورد للتشتت Bradford's law of scattering» لأول مرة على جمهور عريض. وإذا نظرنا بإمعان، فسوف نجد أن كل الأعمال تقريبا التى أسندت إلى كتاب برادفورد تسند أيضا إلى مقالة كتبها ب. س. فيكرى B.C. Vickery: the Journal of documentation, 4,(3), 1948. وحتى لو لم نحصل على عنوان هذه المقالة، فبوسعنا أن نفترض أنها ترتبط بقانون برادفورد بسبب نموذج الاسناد المشترك. وكما طورت دراسة الببليومتريقا، فإن أعمالا أخرى مثل: Human behavior and the principle of the least effort الذى كتبه ج. ك. نيف G.K. Zipf قد أسندت أيضا اسنادا مشتركا. وإن دراسة مثل هذه النماذج أو الأطر هى صعبة فى الوقت نفسه ومرهقة إذا استخدمنا الطرق اليدوية، ولكن كشافات الاسناد المتاحة تجعل ذلك بسيطا نسبيا، ويمكننا أن نتبع الفكرة نفسها خلال مراحل مختلفة^(٨).

إن الكشافات المشتقة مثل كاع SCI لا تحتاج إلى جهد عقلى فى مرحلة المدخلات، حيث أنها مبنية فى الواقع على فرض أن المؤلف قد فعل العمل الضرورى لكى يؤسس لنا روابط الاسناد. وهى تفترض أن المؤلفين يعرفون الانتاج الفكرى لموضوعهم وأنهم سوف يستشهدون بالمصادر المناسبة بصورة كاملة وبشكل صحيح؛ وأنهم لن يتساهلوا فى إسنادات شخصية غير مبررة، ولن يتجاهلوا وثائق تقدم وتبرز آراء متصلة ولكنها معارضة فى الوقت الذى يسجلون فيه مقالات ذات صلة هامشية كتبها أصدقاؤهم. كل هذه الفروض لها ما يبررها إلى حد كبير جدا، ولكن قد لا يكون من الحكمة أن نعتقد أن المؤلفين ليسوا معرضين لأخطاء وهنات الاضافة والحذف مثل أى شخص آخر. ومهما يكن من

شئ، فليس ثمة من شك أن هذه الأدوات هي إضافة مهمة إلى المدى المتاح من الخدمات الببليوجرافية للعاملين في المعلومات. وهي مبنية كذلك على مدخل شائع في البحث عن المعلومات، حيث يبدأ المستفيدون بوثيقة أثارت اهتمامهم والتي يمكن أن تستخدم كبداية للبحث في الكشاف الاسنادى. والمثال الآتى يوضح كيف تعامل مقالة مجلة في الكشاف الاسنادى.

الوثيقة الأصلية تظهر في كشاف المؤلف (المصادر):

Johnson, Karl E. 'IEEE conference publications in libraries', *Library resources and technical services*. 28 (4) October/December 1984, 308-314.
[IEEE = Institution of Electrical and Electronic Engineers]

وهي تحتوى على المراجع الآتية فى النهاية (بين مراجع أخرى):

Marjorie Perego, 'Only the names have been changed to perplex the innocent' *Title varies* 1:13 (April 1974).

Jim E. Cole. 'Conference publications: serials or monographs?' *Library resources & technical services* 22:172 (Spring 1978).

Michlain J. Amir, 'Open letter to IEEE', *Special libraries*. 69:6A (Nov. 1978).

Michael E. Unsworth. 'Treating IEEE conference publications as serials' *Library resources & technical services* 27:221-24 (Apr./June, 1983).

والمداخل الآتية سوف تظهر فى الكشاف الاسنادى:

Amir, Michlain J. 'Open letter to IEEE,' *Special libraries*, 69:6A (Nov. 1978).

Johnson, Karl E. 1984.

Cole, Jim E. 'Conference publications: serials or monographs?' *Library resources & technical services* 22:172 (Spring 1978). **Johnson, Karl E. 1984.**

Perego, Margaret. 'Only the names have been changed to perplex the innocent,' *Title varies* 1:13 (April 1974). **Johnson, Karl E. 1984.**

Unsworth, Michael E. 'Treating IEEE conference publications as serials,' *Library resources & technical services* 27:221-24 (Apr./June, 1983). **Johnson, Karl E. 1984.**

والمداخل الآتية تظهر في كشاف المصطلحات التبادلية Permuterm :

| | | |
|--------------|--------------|------------------------|
| Conference | IEEE | Johnson, Karl E. 1984. |
| Conference | Libraries | Johnson, Karl E. 1984. |
| Conference | Publications | Johnson, Karl E. 1984. |
| IEEE | Conference | Johnson, Karl E. 1984. |
| IEEE | Libraries | Johnson, Karl E. 1984. |
| IEEE | Publications | Johnson, Karl E. 1984. |
| Libraries | Conference | Johnson, Karl E. 1984. |
| Libraries | IEEE | Johnson, Karl E. 1984. |
| Libraries | Publications | Johnson, Karl E. 1984. |
| Publications | Conference | Johnson, Karl E. 1984. |
| Publications | IEEE | Johnson, Karl E. 1984. |
| Publications | Libraries | Johnson, Karl E. 1984. |

والأمثلة السابقة ليس لها بالضبط الاخراج نفسه كما تجدها في : Scienca cita- tion index ، مثلا ، ولكن الأسس هي نفسها. وإذا ما أدخلت البيانات المأخوذة من مقالة المجلة الأصلية، فإن الباقي كله ينتجه الحاسب.

ملخص

تناول هذه الفصل طرق إنتاج الكشافات المطبوعة بواسطة الحاسب من المعلومات الموجودة في وثيقة ما: العنوان أو المراجع البيليوجرافية، والكشافات الاسنادية المذكورة متاحة الآن على الخط المباشر بالتوازي مع النسخ المطبوعة؛ ويصبح اختيار النسخة التي تشتري مسألة اقتصادية وهذا يعتمد على مقدار الاستخدام الذي سوف تتعرض له الخدمة، والكثير من دراسات الاسناد المشترك المذكورة موجودة فقط من الناحية العملية في نسخ على الخط المباشر.

المراجع

- 1 Craven, T. C., *String indexing*, Orlando, Academic Press, 1986. Probably the best text on KWIC KWOC and similar indexes.
- 2 Luhn, H. P., *Keyword in context index for technical literature*, IBM, 1959. Included in *Theory of subject analysis* . . .
- 3 *Chronolog*, 15 (2), 1987, 25, 27 (announcement).
- 4 Bath University Comparative Catalogue Study, *Final report*, Bath University Library, 1975. 10v in 9. (BLR&DD report 5240-5248).
- 5 Garvey, W. D., *Communication: the essence of science*, Oxford, Pergamon, 1979.
- 6 Garfield, E. *Citation indexing: its theory and practice in science, technology and humanities*, New York, Wiley, 1979.
Ellis, P., Hepburn, G. and Oppenheim C., 'Studies on patent citation networks', *Journal of documentation*, 34 (1), 1978, 1-20.
Students should examine at least one of the citation indexes produced by ISI in depth, using it in various ways to test its effectiveness; cf Brahmi, F. A., 'Reference use of *Science citation index*', *Medical reference services quarterly*, 4 (1), 1985, 31-38.
- 7 Mann, T. *Library research models: a guide to classification, cataloging and computers*, New York, NY, Oxford University Press, 1993.
- 8 Small, H., 'Co-citation in the scientific literature: a new measure of the relationship between two documents', *Journal of the American Society for Information Science*, 24 (4), 1973, 265-9; 'Co-citation context analysis and the structure of paradigms', *Journal of documentation*, 36 (3), 1980, 183-96.
Bichteler, J. and Eaton, E. A. III, 'The combined use of bibliographic coupling and co-citation for document retrieval', *Journal of the American Society for Information Science*, 31 (3), 1980, 278-82.
Broadus, R. N., 'Citation analysis', *Advances in librarianship*, 7, 1977, 299-335. (The application of citation analysis to library collection building.)

التطورات في تقانة المعلومات

مع أن هذا الكتاب ليس كتابا في تقانة المعلومات Information technology ، فلكى ندرس استخدام الحاسبات لاسترجاع المعلومات فيجب علينا أن نلقى نظرة على بعض التطورات السريعة في تقانة الحاسبات والتي حدثت خلال السنوات الـ ٣٥ الأخيرة، وبصفة خاصة الـ ١٥ سنة الأخيرة منها، حيث أن لهذه دلالات كبرى، سواء بالنسبة للممارسة الحالية أو بالنسبة للمستقبل، وإن الأساليب التي وصفناها في الفصل ٣ كانت ممكنة مع التقانة المتاحة في ١٩٦٠، ولكن كثيرا من الأساليب التي سوف نغطيها لاحقا لم تصبح ممكنة إلا مع التقانة الأحدث. وبالنسبة لهؤلاء الذين ألفوا هذه التطورات يمكنهم الانتقال مباشرة إلى الفصل ٥ .

جمع الحروف المحكوم بالحاسب

أثناء ستينات القرن العشرين، كانت الحاسبات تجهيزات مركزية كبيرة، والوصول إليها محدود. ولم يكن المستفيد قادرا على التحكم في الناتج النهائي إلا في أضيق الحدود، وكان التشغيل كله تقريبا يتم بالأسلوب المنفصل batch لا على الخط المباشر. وكان إدخال البيانات يتم من خلال بطاقات مثقبة ذات ٨٠ عمودا؛ وكانت الطباعة تتم بسرعة عالية (ولكن بجودة منخفضة) من خلال الطابع السطري. وقد طور جمع الحروف المحكوم بالحاسب Computer - con trolled typesetting فى ذلك الوقت، وقد أنجز قدر كبير من العمل لمؤسسات مثل موط NLM فى مشروعها مدلاز MEDLARS لتحسيب الكشاف

الطبي Index medicus. وإن انتشار تبني هذه التقنية يعني أن جزءا كبيرا من المادة المطبوعة رسميا الآن هي أيضا متاحة في شكل مقروء للحاسب؛ وقبل هذا، فإن أي شيء يشغل بواسطة الحاسب، كان ولا بد أن يثقب بالمفاتيح، وهي عملية مكلفة ومستهلكة للوقت تكرر حقيقة كل العمل الذي تم في إنتاج الوثيقة الأصلية. وإنه بسبب هذا لا تغطي قواعد البيانات البيولوجرافية المادة المطبوعة قبل ستينات القرن العشرين، ولم تصبح مشروعات جعل المواد التاريخية متاحة في شكل مقروء للحاسب ممكنة تقانيا واقتصاديا إلا حديثا.

الحاسبات الصغرى

إن التطورات في تقانة شبه الموصل semiconductor قد أدت إلى إدخال الدوائر المتكاملة والمشغل المصغر micro processor، والذي أدى بدوره في منتصف ستينات القرن العشرين إلى الحاسبات الصغرى الأولى، ولكن كانت هذه الحاسبات «للهواة»، أي: هؤلاء الذين كانوا يستمتعون بالعمل مع الحاسبات ويعرفون معرفة كافية عنها بحيث يمكنهم برمجة الآلات لأنفسهم. ولم يصبح الحاسب المكتبي desktop أو الحاسب الشخصي (حش PC) متاحا للاستعمال العام إلا مع الثمانينات، وسرعان ما تبنته مؤسسات الأعمال والصناعة كما التعليم. ومنذ ذلك الحين أصبحت التطورات في التقانة والبرامج (البرمجيات) تعنى أن أي واحد يمكنه الآن أن يتاح له الآن قوة التحسب نفسها التي كانت متاحة في حاسبات كبرى mainframe منذ عشرين سنة مضت.

ومع الزيادة في قوة التحسب جاءت الحاجة إلى زيادة الذاكرة، سواء بالنسبة للتشغيل أو بالنسبة للتخزين. لقد كانت الحاسبات الشخصية الأولى تملك 64 كيلو بايت (1 كيلوبايت = 1024 (2¹⁰) بايت من رام: ذاكرة توصل عشوائى Random Access Memory (RAM) للتشغيل، واستخدمت الأقراص المرنة Floppy للتخزين الدائم (القرص المرن ذو 360k يمكنه أن يخزن مقابل

١٠٠ صفحة كوارتو). وقد جعلت البرامج الجديدة لتشغيل النظم من الممكن أن تستخدم 640 k من رام، ثم تزيد المقادير؛ والحد الأدنى السائد هو 4 MB (ميغابايت Megabytes)، وتصل آلات أيم IBM المتوافقة Compatible إلى 32 MB. ويمكن للآلات المكتبية ذات المدى العالى Upper range desktop أن تصل قدرتها إلى عدة مئات MB من رام. وكانت الأقراص الصلبة hard disks الأولى، ذات القدرة الكبيرة على التخزين الدائم للملفات تسع 10 MB؛ مثل هذا القرص لايسع الآن برامج تشغيل النظام التى تحتاج إليها إدارة (تنفيذ) الحاسب، والحاسبات التى تسع المئات من الميجايت شائعة الآن.

الأقراص المدمجة (المليزرة) CD - ROM

إن الأقراص الصلبة هى فى العادة جزء دائم من الآلة - ومن ثم كان اسمها أولا القرص الثابت Fixed - ولكن تطوير القرص المدمج المقروء بالليزر قد وفر إمكانية التخزين حتى 680 MB (مقدرة ١٩٩٥) من البيانات على قرص مدمج CD - ROM متحرك (Compact Disc - Read Only Memory). والقرص المدمج CD - ROM هو الشكل الموجود عادة، رغم أنه ليس النوع الوحيد من القرص المقروء بالليزر. وهناك شكل ظهر قبل القرص المدمج هو قرص الليزر ١٢ بوصة؛ وهذا وسيلة (نيطة) تناظرية analog device، تسجل البيانات فيه فى شكل مطابق لمقاييس التلفزيون، وليس الشكل الرقى المقروء للحاسب من قم CD. وقد استخدم بنجاح لتخزين الإيضاحيات، مثال ذلك الصور فى مجموعات الأرشيف. وكما يعنى الاسم، فإن الأقراص المدمجة هى وسائل قراءة فقط، ولكنها يمكن أن تغير فيما بينها بسرعة بحيث توفر الإتاحة لمصادر متنوعة من المعلومات، وإن تبنى معيار دولى (أيزو ٩٦٦٠ 9660 ISO) - الذى يعرف أيضا باسم (the High Sierra standard) قد أدى إلى زيادة سريعة فى عدد قواعد البيانات المتاحة على قم. والتى تغطى مدى واسعا من المعلومات.

ويعنى التوحيد القياسى أن الأقراص من أى مورد يمكن أن تقرأ على أى مسير رقم CD - ROM drive : فليس السوق مقصورا بأى حال على الناس الذين لديهم تجهيزات المضاهاة .

ويمكن لقرص مدمج واحد أن يتسع لتخزين نص دائرة معارف كاملة، مع الايضاحيات - التى تضم الفيديو كليب - والصوت . وربما كان الأهم حاليا هو إمكانية تخزين قواعد بيانات بيلوجرافية كاملة، تضم النصوص الكاملة، بحيث لا نكون مكرهين بالحاجة إلى ربطها بحاسب مركزى، ولكن يمكننا أن ننفذ البحوث على حاسباتنا الشخصية . وبالنسبة للكثيرين من المستخدمين، فهذا ليس مريحا فقط ولكنه أقل ضغطا ! (فهو أيضا يتجنب تكاليف الاتصالات من بعد) . وسوف نناقش بعض دلالات جوانب التقدم التقانى هذه هنا، ولكننا لايمكن أن نأمل فى تغطيتها بعمق؛ وإن مزيدا من القراءة هى أمر ضرورى للحصول على تقدير كامل للامكانات .

الشبكات

مالم نكن قادرين على ربط حاسبنا الخاص بالحاسبات الأخرى، فسوف نكون مقتصرين على قواعد البيانات التى توجد على آلتنا أو على واسطة يمكن نقلها مثل القرص المدمج . وإن استخدام قواعد بيانات خارجية يحتاج منا أن نكون قادرين على ربط الحاسبات معا . وفى الأيام الأولى للربط البينى (الداخلى) interconnection، كانت الحاسبات الكبرى main frame متاحة لمضاعفة المستخدمين عن طريق روابط مباشرة لطرفيات غير ذكية: وهى غير ذكية dumb من حيث أنها لايمكن أن تعمل بصورة منفصلة عن الحاسب الكبير . أما تطوير الحاسب المصغر microcomputer فإنه يعنى أن هذه يمكن ربطها بحاسبات أخرى، ولكنها تعمل كطرفيات ذكية، بمعنى أنها تنفذ التشغيل بنفسها . وقد كان الأكثر معقولة الاستفادة من قوة الآلة المحلية حتى ولو ارتبطت

بحاسب مركزى كبير. ولكى يحدث هذا، كان من الضرورى وجود أمرين:
روابط الاتصالات عن بعد، وبرامج مناسبة.

البرامج (البرمجيات)

لقد طورت شبكات المناطق المحلية (لان LAN) Local Area Networks باستخدام البرامج التى ربطت الحاسبات الشخصية بحاسبات شخصية أخرى وكذلك بحاسبات كبرى فى نطاق المؤسسة نفسها. وفى الوقت نفسه، كان المستفيدون العسكريون والأكاديميون والتجارىون ينفذون فكرة ربط حاسبات فى مواقع بعيدة من خلال شبكات المناطق الواسعة (وان WAN) Wide Area Net-works. وفى سنة ١٩٦٩، أسست وزارة الدفاع الأمريكية شبكة أربانت ARPANET، وكان غرضها منع الاضطراب الكامل لشبكة الدفاع فى حالة التعرض لهجوم نووى، من خلال توزيع قوة التحسب على عدد من المواقع Sites المتباعدة. وفى منتصف الثمانينات رأت المؤسسة الوطنية للعلوم (موج NSF) إمكانات استخدام هذا الأسلوب لتخفيض تكلفة البحث فى الحاسبات الفائقة Supercomputers؛ بدلا من أن يكون لكل مركز بحث أكاديمى حاسبه الفائق الخاص، والذي قد لا يستفاد منه استفادة كاملة، ارتبط عدد محدود من المراكز التى تمتلك حاسبات فائقة معا من خلال موجت NSFNET وأصبحت تكلفة الوصول إلى المستفيدين الآخرين فعالة إلى درجة أكبر بكثير مما سبق. وقد تأسست شبكات أكاديمية أخرى مثال ذلك: JANET جانيت^(١) فى المملكة المتحدة، يورن UREN فى وما USA، وآرنت AARNET والشبكة الأسترالية التعليمية (ed NA) فى استراليا؛ ولما زادت الشبكات التى ترتبط معا، أصبحت أساس الإنترنت Internet، شبكة الشبكات التى تربط ملايين المستفيدين فى أقطار العالم. وقد اقتنعت المؤسسات التجارية مثل البنوك وشركات الطيران بسرعة بقيمة خدمات شبكات الحاسبات التى تربط الفروع فى

المدينة نفسها، والقطر نفسه، ودوليا، وطورت برامج معقدة إلى درجة عالية جدا لتسهيل أعمال البنوك والسفر على مستوى العالم. وعلى عكس الانترنت، فإن هذه الشبكات، وكذلك تلك التي طورت لأغراض دفاعية، ليست للاستعمال العام، رغم أن بعض قراصنة الحاسب hackers قد خرقوا هذا، ولتسهيل الاستفادة من المعلومات على حاسب بواسطة حاسب آخر بعيد، طور مفهوم برنامج وحدة خدمة العملاء (الزبائن) Client - Server. والحاسب الذى يقدم المعلومات هو وحدة خدمة Server، وتصمم البرامج لتقديم المعلومات فى شكل مناسب للحاسبات الأخرى، ربما لأكثر من واحد فى الوقت نفسه. والحاسب الذى يستفيد هو العميل Client، الذى يجب مرة أخرى أن يكون عنده البرامج الصحيحة (المناسبة) للاستفادة من المعلومات فى الشكل الذى تقدمها عليه وحدة الخدمة. فإذا كنا نتحدث عن النقل البسيط للنص، فهذا ليس أمرا صعبا، ويوجد هنا فارق طفيف بين العميل وبين وحدة الخدمة؛ وسوف نناقش فيما بعد الكود الموحد لنقل النص آسكى ASCII. ويمكن أن تنقل المؤلفات كذلك فى شكل ثنائى؛ وهذا هو الشكل المطلوب لملفات الوسائط المتعددة، التى تكون كبيرة وللإستفادة من هذه الملفات، فيجب أن يكون لدى المستفيدين البرامج المناسبة علي حاسبهم الشخصى، بالإضافة إلى البرامج المطلوبة للشبكة.

الاتصالات عن بعد

لقد كانت التطورات فى مجال وسائل الاتصال عن بعد telecommunication links عاملا رئيسيا فى نمو التحسب الموزع. فى البدء، استخدمت شبكة الهاتف الموجودة؛ ولما كانت هذه الشبكة تستخدم تجهيزات ميكانيكية كهربائية، فقد كانت عرضة لمستويات عالية من الخطأ، وكانت معدلات النقل بطيئة - كانت أصلا 110 bps (bits per second) والتي سرعان ما ارتفعت إلى 300 bps. (عدد البتات bits المنقولة فى الثانية غالباً ما يطلق عليه الاسم معدل بود: معدل نقل

الوحدات الثنائية baud rate ، ولكن عند السرعات الأعلى يصبح من الشائع أن الإثنين لا يكونان متساويين equivalent أبداً). وإن إحدى المشكلات هي أن الحاسبات تنتج إشارات Signals قيمة تتألف من سلسلة من الأصفار وأعداد ١؛ أما دوائر الهاتف القياسية فهي وسائل (نبائط) تناظرية analogue devices ، تجسد الصوت بواسطة فولت كهربائي متغير باستمرار. ولتحويل الإشارات الرقمية إلى شكل يمكن نقله عبر أسلاك الهاتف فهذا يحتاج إلى مودم (معدّل) - (modulator) modem (demodulator) الذى يحول الاشارات من رقمية إلى مثلتها analogue عند إحدى النهايات، ومن مثلتها إلى رقمية عند الأخرى. ولا تقتصر المشكلة على احتمال الضوضاء أثناء عملية الاتصال عن بعد، بل إن هناك أيضا احتمال الخطأ فى عملية التحويل. وفى السنوات الأخيرة حلت الوسائل الالكترونية محل الوسائل الميكانيكية الكهربائية القديمة، وهناك تحسينات فى البرامج والأجهزة الخاصة بالمودمات (المعدلات)، وهكذا أصبح الحد فى الوقت الحاضر (١٩٩٥) هو ٢٨,٨٠٠ بتات ثانية، وهو حد مقبول دوليا (V 34) فى ١٩٩٤. وإذا قارنا هذا بالسرعة التى يعمل بها الحاسب حاليا، فإنه لايزال بطيئا جدا، وتبذل جهود كبيرة لزيادة سرعات النقل. وأحد الأساليب لذلك هو زيادة فاعلية استخدام شبكة الهاتف الموجودة، الأمر الذى يحتاج إلى استثمارات ضخمة فى البنية التحتية. وإن تحويل الحزمات Packet Switching، الذى بدأ فى سبعينات القرن ٢٠، يعنى أنه قد أصبح من الممكن ربط شبكات عبر مسافات بعيدة بتكلفة معقولة. وإن الاتصال الهاتفى المعتاد يحتكر monopolizes جزءا معينا من الروابط المتاحة بين المركزين المعنيين - عرض النطاق الترددى bandwidth -، ولكن كم المعلومات المنقولة يشغل جزءا محدودا فقط من الوقت. وعلى سبيل المثال، لا يرسل فيما بين الكلمات أى شىء فى الاتجاهين. ويمكن للحاسبات أن تقوم بقدر كبير من التشغيل أثناء تلك الفترات الميتة، مثلما تستطيع أن تفعل بين ضربات المفاتيح (الحروف). ويأخذ تحويل الحزمات المدخلات من عدد من الرسائل ويقسمها إلى «حزمات»،

مضغوطة مميزة "packets" labelled compressed؛ ثم ترسل الحزمات إلى محطة الاستقبال في تيار مستمر (متدفق)، حيث تخزن في شكل الرسائل الأصلية، المرسلة إلى المستقبلين المعنيين. والوقت الميت يضغط، ويمكن إرسال مرور أكبر جدا عبر قناة النقل نفسها باستخدام بروتوكول X. 25.

وفي منتصف الثمانينات قدم اقتراح بأنه يمكن استخدام شبكة الهاتف لنقل الاشارات الرقمية في Inregrated Services Digital Network (الشبكة الرقمية للخدمات المتكاملة شرحم ISDN). وهذه الشبكة لايمكنها فقط نقل المكالمات الهاتفية وأن تربط الحاسبات، ولكنها يمكن أن تقدم كذلك الوسائل لتسليم إشارات الفيديو، بسرعة 128 Kbps أو أسرع. والتغيير إلى شرحم ISDM لايزال بطيئا في الوقت الحاضر، ولكن يبدو من المرجح أن المعايير القياسية الدولية الجديدة والهبوط السريع في التكلفة سوف يغير هذا بسرعة. ومع استعمال ISDN شرحم لا يكون هناك طبعاً حاجة إلى مودم في أى نهاية من نهايات الحاسب لكي يكون قادراً على الاتصال، طالما أن المخرجات والمدخلات الرقمية يمكن نقلها مباشرة دونما حاجة إلى تحويلها في شكل مماثل أو مناظر.

وتتألف الأسلاك التقليدية للهاتف من أزواج من الأسلاك النحاسية (تعرف بإسم «الأزواج الملتفة twisted pairs»); وكمية المعلومات التي يمكن لهذه الأسلاك حملها كمية محدودة. وقد كان أحد التطورات المهمة إدخال كابلات الألياف البصرية Fibre - optic cable. وتحول الإشارات الكهربائية إلى نبضات من الضوء وتنقل عبر ليفة زجاجية طويلة دقيقة جداً؛ وفي النهاية البعيدة تحول مرة ثانية إلى إشارات كهربائية. ويتم النقل بسرعات عالية جداً: حالياً 50 Mbps، ومن المحتمل أن سرعات $2 \text{ G} (10^9)$ سوف تكون ممكنة في وقت قريب. وفي

الدول المتقدمة، يتم الآن تغيير روابط الهاتف داخل المدن إلى كوابل الألياف البصرية. وإن الزيادة الضخمة في المقدرة على الحمل تعنى أن كابلا واحدا يمكن أن يحمل العدد نفسه من الإشارات الذى يحمله عدد كبير جدا من روابط الأسلاك النحاسية الزوجية الملتفة. وإن كابلات الألياف البصرية هى وسيلة عملية لحمل الإشارات التليفزيونية كما الصوت، مثلا. وإن الشبكة النامية من كابلات الألياف البصرية هى البنية الرئيسية لطرق المعلومات السريعة الفائقة information super highway، التى نسمع عنها كثيرا. ومثلما يمكن للطرق الحرة أن تحمل مرورا أكثر بسرعات عالية من شبكة الطرق المحلية، كذلك فإن شبكة الألياف البصرية يمكنها أن تحمل معلومات أكثر من شبكة الهاتف التقليدية.

والنوع الثالث من الروابط يستخدم النقل عبر الميكروويف. وهذه تحتاج إلى هوائى طبق عند كل نهاية للرابطة، ولكنها لا تحتاج إلى توصيلات أو وصلات Connections أخرى. وتستخدم هذه الطريقة لنقل البيانات، بطريقة خطوط الهاتف نفسها. ويحد منها أنه لا بد أن يكون هناك نقل - of - sight - line transmission من نقطة للنقطة التى تليها، ولكن إذا تحقق هذا، فإن التكاليف يمكن مقارنتها مع الكابل. وإن استخدام الأقمار الصناعية يعنى أن الإرسال عبر الميكروويف (مثل ذلك TV) يمكن أن يذاع أيضا. أى يوزع على مستوى العالم، وإن المستقبلات الهوائية للأطباق هى الآن منتشرة وشائعة (المعلومات من السماء!). ويجرى الآن إلى حد كبير حل مشكلات ربط الحاسبات فى الشبكات، رغم أن البحث فى زيادة سرعات النقل (الإرسال) سوف يستمر بالتأكيد. وهكذا فإنه يوجد تحت تصرفنا الوسائل المادية للوصول إلى المعلومات المخزنة فى مدى هائل من الحاسبات. وعلينا الآن أن ندرس كيف تقدم هذه المعلومات.

المواجهة بين الرسوم والمستفيد

إن أحد التطورات المهمة فى عالم الحاسبات الصغرى هو دخول آبل ماكنتوش Apple Macintosh فى سنة ١٩٨٤ . وقبل هذا كان لابد من كتابة أوامر الحاسب، إما من الذاكرة أو من خلال البحث فى دليل . ولسنوات عدة، كان هذا نشاطا مرهقا ومسيبا للخطأ، كما جعل استخدام الحاسبات الشخصية يقتصر على من يملكون المهارات اللازمة . وقد كان آبل ماكنتوش أول حاسب صغير يستخدم: مواجهة الرسوم والمستفيد مرم graphical vsev interfoce, GUI الذى تمثل فيه الأوامر على الشاشة بواسطة أيقونات icons، تختار بواسطة وسيلة «إشارة» مثل الفأر . وقد استخدم الفأر لكى يحرك الشعرة Cursor إلى الموضع المطلوب على الشاشة؛ ثم يتم إصدار الأمر المطلوب عن طريق طقطقة Click واحدة أو اثنتين من زر أو أزرار الفأر . وفى المقابل فإن القوائم - قوائم الأوامر - يمكن اختيارها بالطريقة نفسها . وقد قللت كتابة الأوامر إلى الحد الأدنى . وهذا يجعل الآلة أبسط فى الاستخدام . وبالإضافة إلى ذلك، فإن استخدام الرسوم فى داخل البرامج قد أصبح أسهل من خلال حقيقة أن عرض الشاشة كلها عبارة عن رسم؛ وإن إضافة الايضاحيات المطلوبة هى برمجة أسهل وأقل تعقيدا من الشاشات التى تعرض النصوص . وقد احتاج إنتاج مرم GUI ذى جودة مساوية فى الحاسبات الشخصية الموافقة له أبم إلى الانتظار حتى وندوز ٣,١ فى سنة ١٩٩١ .

وسائل المدخلات

بالنسبة للنص، فلازالت لوحة المفاتيح هى الوسيلة القياسية للإدخال وتشبه لوحة المفاتيح فى الأصل وإلى حد كبير جدا لوحة مفاتيح الآلة الكاتبة مع بعض المفاتيح الزائدة . وبالرغم من بعض التغييرات، فإن الشكل القياسى قد أحسن عمله لدرجة أنه من المنتظر أن يستمر، مثلما استمرت لوحة مفاتيح

الآلة الكاتبة. وهناك وسيلة أثبتت فائدتها في تحويل النص إلى شكل رقمي وهي الماسح (الباحث) Scanner، الذي يحول أيضا الرسوم بما فيها الملونة. ومعنى هذا أن المكتبات الكاملة التي تحتوى على الرسوم فقط يمكن تجسيدها في الوثائق حسب الحاجة. والأحدث من هذا استخدام كاميرات الفيديو الرقمية، التي تمكننا من أن نضيف الصور أو الفيديو كليب إلى وثائقنا، وبالألوان. وهناك وسيلة أخرى من المؤكد أن أهميتها سوف تزيد وهي المدخلات الصوتية؛ إن الوسائط المتعددة للحاسبات الشخصية لديها بالفعل القدرة على إخراج الصوت من النص، ولكن إدخال الصوت أيضا يصبح عمليا، رغم أنه حتى الآن لا يزال تجريبيا^(٢).

النشر المكتبي

إن دخول طابعة الليزر في ١٩٨٥ يعني أن مخرجات الحاسب لم تعد محدودة بحدود الطابعات الأقل تعقيدا. ويمكن استخدام طواقم fonts مختلفة من حروف الطباعة، مع رسوم ذات جودة عالية تجسد في النص، وسرعان ما تطورت البرامج للاستفادة من هذه الامكانيات، بما في ذلك الطباعة بالألوان. وكثيرون من الناس الآن ينتجون نشراتهم وبحوثهم الخاصة. ومع أن هذه جميعا تنتج بواسطة الحاسب، فإنها لا تدخل ضمن قنوات النشر المعتاد لكي تلتقطها المؤسسات البيولوجرافية؛ هؤلاء الذين لا يصبحون جزءا من الأدب «الرمادي»^(٣)، سريع التزايد، ولا يمكن اقتفاء أثرهم إلا عن طريق الصدفة وحدها. والوجه الآخر للعملة هو أن الأعمال المفيدة، مثل بعض المطبوعات المتخصصة التي تصدر منها طبعات صغيرة فقط، يمكن إنتاجها بهذه الطريقة وطباعتها عند الطلب. وتمكن طابعات الليزر الحديثة العالية السرعة وآلات التجليد - تمكن حوانيت الكتب أو الناشر المكتبي من أن يحتفظ بالنص في شكل مقروء للحاسب، وتنتج نسخة «أثناء انتظارك». وقد جعل جمع الحروف المحكوم بالحاسب معظم المادة المطبوعة بصورة رسمية مقروءة بواسطة

الحاسب؛ وقد فعل النشر المكتبي الشيء نفسه بالنسبة للمطبوعات غير الرسمية^(٤).

النشر الإلكتروني

إن واحدة من النتائج المهمة هي تطوير النشر الإلكتروني - electronic publishing. وتنشر الآن بعض المواد على الخط المباشر، وهي تشمل على مجلات، وتوزع على الإنترنت^(٥). وبسبب أنه لا توجد طريقة لدفع قيمة المادة التي تنشر بهذه الطريقة، فقد أصبحت مناسبة لمطبوعات مثل المجلات الأكاديمية، تقارير الطقس والوثائق الحكومية التي لا يقصد منها الربح. ولما بدأ التعامل مع الإنترنت على أساس تجارى، يصبح دفع القيمة عمليا، وهناك مجلات أخرى كثيرة تصبح الآن متاحة بهذه الطريقة.

المكتب اللاورقى

فى مؤسسات كثيرة الآن يكون لكل عضو هيئة تدريس حاسب مكتبي يكون جزءا من شبكة المنطقة المحلية (لان LAN). وليس من المحتم أن توزع المعلومات فى شكل مطبوع على الورق، ولكن يمكن توزيعها بالوسائل الالكترونية باستخدام برامج مناسبة^(٦). ويمكن للمستفيدين الأفراد بناء ملفاتهم الشخصية الخاصة للمعلوما المفيدة لهم، وهذه المعلومات يمكن نقلها باستخدام البريد الالكتروني إلى شخص واحد، أو إلى جماعة مختارة، أو إلى المؤسسة ككل، أو إلى أى شخص يريد أن يقرأها (يجب اتباع الاجراءات الأمنية لضمان أن المعلومات لن تصل إلا إلى الأشخاص المخولين بقراءتها!) والمؤسسات الكبيرة مثل CIA, ICI كانت أول من طبق هذه الطريقة للعمل، ولكنها حتى الآن لم تصبح بعد الدواء العام لكل الأمراض والذي افترض يوما ما، ومعظم المكاتب لاتزال مبنية على الورق. (فى بعض الحالات، هناك ضرورات قانونية تتطلب أن تكون الوثائق على الورق). ومن الممكن أن تنقل رسائل الفاكس من حاسب إلى حاسب، ولكن آلة الفاكس بنسخها الورقية لازالت هى الأساس،

حتى رغم أن معظم الرسائل تنتج بواسطة مشغل (معالج) الكلمات world pro-cessor وتطبع بغرض النقل.

وقد كانت Bulletin boards متاحة لبعض السنوات، وتتألف هذه من حاسبات يمكن الوصول إليها بواسطة مودم وخط هاتف. وبعضها من عمل هواة متحمسين، وبعضها أسسته مؤسسات من نوع أو آخر. ويمكن أن توضع الرسائل على اللوحة بواسطة SysOP (System Operator) الذي يدير the BBS، أو بواسطة المستخدمين الذين يرتبطون باللوحة من خلال حاسباتهم الشخصية الخاصة. وبعضها متاح بصورة مجانية، بما في ذلك البعض الذي يعتبر مصادر رئيسية للملكية العامة أو برامج sharware، ولكن بعضها الآخر، مثل Compu-serve، تحصل على مقابل مادي ولكنها تقدم مدى أوسع من المعلومات والخدمات. وإن توزيع المعلومات بواسطة القنوات الالكترونية هو الآن أمر منتشر.

النص الفائق

حينما نقرأ قصة، فإننا نتوقع أن نقرأها مرة واحدة دون توقف، لكي نتابع تطور الأحداث والأشخاص. وفي المقابل، لو كنا نقرأ شيئاً ما في دائرة معارف، فإننا نرغب في متابعة إحالة لمقالة في مكان آخر في العمل، ربما في مجلد آخر؛ وربما أُحِلْنَا من هناك إلى مقال آخر. وهكذا فإن صفوف الأفكار ليست خطية أو طولية، كما أشار إلى ذلك فانيفار بوش Vannevar Bush^(٧). في مقال مفتاحي. إن فكرة أن هذا النوع من الاستطلاع أو الانتقال من رسالة إلى أخرى يمكن عمله بواسطة الحاسب قد اقترحها لأول مرة تيد نلسون Ted Nelson^(٨)، ولكن أول تطبيق عملي على نطاق واسع جاء مع دخول البطاقة الفائقة Hypercard على آبل ماكنتوش. لقد مكنت هذه المستفيد من أن يجمع ملفاً File (رصّة stack) مع روابط مبنية داخلياً Builtin خارج التسلسل الرئيسي. وكان الاسم الذي أعطاه نلسون لهذا النوع من بنية الملفات هو النص الفائق hyper-text، من فكرة قدمت كمفهوم رياضي في القرن ١٩، وهي الحيز الفائق hyper-

space المتعدد الأبعاد. والنص الفائق متاح الآن بصورة عامة وهو مستخدم نطاق واسع في قواعد البيانات على أقراص مدمجة، على سبيل المثال.

تبدأ روابط النص الفائق من نقطة البداية في الرابطة، المرتكز the anchor. وحينئذ يكون من الضروري تخصيص الموضوع الدقيق الذي ينبغي أن يربط هذا معه، وهذا يصنع رابطة فائقة مع مرتكز آخر. فإذا ما تم هذا، فإن الموضوع الجديد يمكن أن يصبح مرتكز نقطة بداية، مع توليد أو تأصيل روابط أخرى عند الحاجة. قد تكون الروابط مخصصة، تأخذ المستفيد إلى موضع أو مكان مخصص؛ وقد تكون محلية، تأخذ المستفيد إلى أى نقطة مختارة في الوثيقة الجارية؛ أو شاملة، تأخذ المستفيد إلى أية نقطة فى أية وثيقة - والتي قد تكون على حاسب مستقل تماماً، والذي بدوره قد يقود إلى ملف على حاسب آخر أيضا. ومن السهل غالبا أن نفقد المسار الذى وصل إليه الشخص!

الوسائط المتعددة والوسائط الفائقة

إن مرم GUI قادر على عرض الرسوم مثلما هو قادر على عرض النصوص. ومع النمو فى قوة الحاسب، أصبحت هناك ذاكرة كافية متاحة لكى تجعل هذا ممكنا بتكلفة معقولة؛ وتشغل ملفات الرسوم حيزا من الأقراص أكبر بكثير جدا من ملفات النصوص. ولم تصبح الرسوم اقتراحا عمليا حتى طور تشغيل 32-bit وليس 8-bit أو 16-bit، وتم هذا التطور مع آبل ماكنتوش، وقد تبعه بعد سنوات the intel 80386 والمعالجات (المشغلات) Processors الأحداث من نوع حاسبات آيم IBM الشخصية. ويمكن لحاسبات الناشر المكتبى التى تستخدم بصفة خاصة كمحطات عمل للرسوم - يمكنها أن تستخدم تشغيل 64-bit. وإن التغيير من تشغيل 8-bit إلى 32-bit قد صاحبه زيادات فى السرعة التي يمكن أن تعمل بها المعالجات (المشغلات) الدقيقة microprocessors؛ فى حين أن المعالج القديم الذى يستخدم 9 بتات يعمل عند $1 \text{ cycle per sec} = 1 \text{ Hertz}$ (1 Hertz) 4MHz (ond)، أو 10MHz فى نوع التريبو، وتصل موديلات 1995 حتى 130MHz. هذه السرعات العالية أساسية لتشغيل الرسوم، وبخاصة الفيديو.

إن التسجيل الرقمي للصوت، والذي بدأ على تسجيلات LP ثم على قم CD معناه أن الصوت أيضا يمكن تجسيده في ملفات الحاسب. وتشغل ملفات الصوت أيضا حيزا : الوتر chord الذى يقدم ويندوز يشغل حوالى 25k للثانية أو مثلها لوقت التشغيل playing. وإن دخول الرسوم (الصورة) والصوت فى ملفات الحاسب أدى إلى ظهور العروض المتعددة الوسائط multimedia presentations، التى يمكن الآن أن تشمل الفيديو والرسوم المتحركة animations. وانطلاقا من هذا فقد كانت الخطوة التالية هي الوسائط الفائقة hypermedia : الوسائط المتعددة التى يمكننا فيها أن نقفز من نقطة لأخرى عبر روابط النصوص الفائقة^(٩).

تبادل البيانات

إن أحد أوجه القصور المهمة للشاشة التى تعرض النصوص هى حقيقة أنها لايمكن أن تعرض إلا عددا محدودا من الرموز : الحروف الكبيرة والصغيرة، والأرقام وعلامات الترقيم. وقد طورت كودات لتمثل هذه الرموز فى كود ثنائى binary ، ولكن الكود المقبول الآن على نطاق واسع هو مجموعة حروف الآسكى ASCII (American Standard Code for Information Interchange) ^(١٠). وإن تسلسل الكودات مهم، لأنه يحدد نظام التصنيف للرموز المختلفة. وهكذا فإن استخدام كودات آسكى يمكنه أن يملئ نظام التصنيف الذى يوجد فى قوائم رؤوس الموضوعات المبنية على الحاسب.

ويستخدم هذا المعيار الموحد على نطاق واسع جدا، ولكنه يصلح فقط للألفباء اللاتينية، وليس للحركات accents والحروف الخاصة التى تستخدم فى اللغات الأوربية. ويوجد الآن مجموعة آسكى موسعة ٨ بتات، ولكنها لا تغطي كل الحاجات بصورة مناسبة حتى بالنسبة للغات الأوربية. ومن الواضح أن تبادل المعلومات سيكون محدودا للغاية إذا اقتصرنا على مجموعة حروف آسكى،

ومع ذلك فلا زال هذا هو الكود القياسى المستخدم للبريد الإلكتروني، ببساطة لأنه قياسى! وليس هناك ضمان بأن المرسل والمستقبل لرسالة البريد الإلكتروني يستخدمون الرموز الموسعة نفسها، ولكن إذا قصرنا نصهم على أسكى القياسى ٧ بتات فسوف تستقبل كما أرسلت. وإن لغات كاليونانية التى لاتستخدم الألفباء الرومانية قد تضطر إلى تبنى التوافق المعقدة لكودات أسكى.

وهناك حد قاس بالنسبة للرسالة الأيسط هى أنه لايمكن نقل معلومات عن الصياغة formatting. ويمكننا معالج (مشغل) الكلمات من استخدام طواقم حروف طباعة متنوعة، وطواقم أحجام، وطواقم التأكيد (البنط الثقيل bold، الحروف المائلة italic، وضع خط) وطواقم إخراج، وذلك لإنتاج وثائق تبدو حسنة وتحمل المعنى بصورة فعالة. ولايمكن نقل أى من هذه باستخدام أسكى. وقبل دخول مرم GUI، لم يكن هذا حيويا، حيث أن التأثيرات لم يكن من الممكن عرضها، إلا بصورة مباشرة؛ ومع عرض GUI، فإن كل هذه الآثار الخاصة يمكن عرضها على الشاشة - ولكن لازال من غير الممكن نقلها إلى حاسبات أخرى تستخدم كودات أسكى. من الممكن فك شفرتها فى شكل ثنائى ثم نقلها - معظم برامج الاتصالات تنقل الملفات الثنائية - ولكنها لن يكون لها أى معنى فى طرف الاستقبال إلا إذا استخدم معالج الكلمات نفسه بالضبط لعرضها. ومن ناحية أخرى، فإن القدرة على عرض شكل على مرم GUI قد يقودنا إلى إهمال البنية والمحتوى لصالح الإخراج؛ ويحتاج هونان Hon-an^(١١). أنه بالنسبة للوثائق الكبيرة يكون معالج الكلمات المبنى على النص فعلا بالدرجة نفسها مثل معالج مبنى على مرم GUI، طالما أننا مضطرون إلى الاهتمام بالمحتوى لا أن نسمح لأنفسنا بأن نكون مشغولين أكثر من اللازم بالشكل الذى تقدم عيه.

المعايير القياسية

إن كلا من الأجهزة والبرامج عرضة للتغيير، وغالبا ما يكون التغيير سريعا

جدا. وقد كانت هناك حاجة إلى تطوير معايير قياسية Standards تجعل من الممكن تبادل البيانات بصرف النظر عن الأجهزة والبرامج المستخدمة لإنتاج هذه البيانات^(١٢). وكانت الخطوة الأولى تطوير SGML: Standardized General Markup Language في سنة ١٩٨٦^(١٣). وتعمل هذه عن طريق وضع تاج لكل وحدة من الوثيقة، مثل الرأس، العنوان، النص، حتى يتسنى تعرف المستقبل عليها كما هي. ن كذلك يمكن فك شفرة المعلومات المتعلقة بالاعراج والأبناط. والكود الواسع الانتشار HTML : Hyper Text Markup Language هو مجموعة فرعية من SGML. وهو يوفر تاجات للرؤس، والعنوان وما إلى ذلك. وكل تاج يجب أن يبدأ وينتهي. والنتيجة النهائية هي وثيقة تستخدم كودات آسكى القياسية ولذلك يمكن نقلها ببساطة إلى حاسب آخر، حيث يمكن فك شفرتها بواسطة برنامج استطلاعى مناسب. وإن صفحة بسيطة يمكن أن تظهر كما يأتى :

```
<html>
<head>
  <title>Welcome</title>           [headings]
  <h1>Welcome!</h1>
</head>
<body>
  [body tags and text]
</body>
</html>
```

كل تاج تبدأ أو تنتهى وذلك حتى يتسنى للمستقبل أن يتعرف على مختلف أجزاء الوثيقة. وتبين النهاية بشرطة مائلة تسبق التاج^(١٤). وبالنسبة لهؤلاء الذين لا يرغبون فى أن يقوموا بأنفسهم بعملية الكويد، فإن هناك برنامجا يقوم بها أتوماتيا، أو (عند طرف المستقبل)، يجرد الكودات ويعطى نص آسكى بوضوح.

وقد طورت معايير قياسية أخرى بما فى ذلك

وهناك مشروع مهم وهو Text Encoding Initiative (TEI)، الذى طور بالمملكة المتحدة (مم UK) لكى يسمح بتبادل الوثائق بين الجامعات^(١٥). وربما كان إرسال نص بواسطة البريد الالكترونى باعتباره ملف آسكى أسرع، ولكن هذا يعوقه كمية العمل التى يتضمنها إعادة تشكيل الوثيقة بحيث تكون متوافقة مع الشكل وقد استخدم TEI شكلا من SGML لحل المشكلة وهناك حل آخر ممكن هو استخدام برامج التشفير / فك الشفرة : UUENCODE. EXE و UUDCODE. EXE؛ وهذه البرامج تحول ملفات الحاسب إلى خيوط آسكى التى يمكن نقلها عبر الانترنت وفك شفرتها فى نهاية الاستقبال.

إن دخول الوسائط المتعددة والوسائط الفائقة يعنى أن من الضرورى تطوير معايير قياسية جديدة، لا لمجرد تغطية كل شكل على حدة فحسب ولكن لاستعمالها معا كذلك، طالما أن من الضرورى أن تكون متزامنة Synchronized (in sync). ويمكن أن يكون الأمر محبطا disconcerting إذا لم يكن الصوت والصورة متزامنين! وإن واحدا من المعايير القياسية المفاتيح هو Hy Time، وهو لغة بنوية structuring مبنية على SGML وسائط متعددة / مبنية على الوقت : SGML - based Hypermedia / Time-based Structuring lanague؛ وتخدم الويقة الأصلية التى حلت شفرتها كمحور hub لملفات النص، وملفات الصوت، والرسوم، وتربطه جميعا وتضمن أن البيانات من كل منها تستخدم فى اللحظة المناسبة.

وهناك مظهر آخر مهم لعملية نقل الملفات وهو كمية المعلومات التى تنقل. ويعرض مراقب monitor الحاسب المعلومات على شكل صور Pixels، ونقط مفردة للون؛ ويحتوى التحليل الشائع الذى يضم ٦٤٠ x ٤٨٠ صورة pixels على ٣٠٧٢٠٠ صورة. وبالنسبة للعروض ذات ٢٥٦ لونا، تحتاج كل صورة إلى ٨ بتات (٢٨ = ٢٥٦)، بإجمالى قدره ٢٤٥٧٦٠٠ بت من المعلومات،

أو ٣٠٧٢٠٠ bytes. وعلى هذا فإن صورة ثابتة static في ٢٥٦ لونا تشغل مايزيد على 300KB من حيز القرص. ويستخدم ماكتوش لونا ذا ١٦ بت، وهذا يعطى ٦٥٠٥٣٦ لونا، ويحتاج إلى زيادة الحيز بمقدار الضعف.

ولكى تعرض الصور المتحركة، فمن الضروري أن تكرر الصورة ٦٠ مرة في الثانية لكي تستفيد من ظاهرة استمرار الرؤية في نطاق بارامترات شاشة الحاسب. ومن الواضح أن كمية المعلومات التي يجب تخزينها كبيرة جدا، وقد طورت أشكال الضغط لتقليل هذه الكمية إلى مستوى مقبول. وهنا أيضا نجد أن المعايير القياسية مهمة جدا؛ لاستخدام ملف رسوم مضغوط فلا بد أن يكون لدينا البرامج المناسبة لإبطال الضغط decompress. وإذا كنا ننقل ملف رسوم عبر شبكة ما باستخدام مودم (معدل)، فمن الواضح أنه حتى عند 28.8kbps فإن الأمر سوف يحتاج إلى بعض الوقت لنقل ملف يتألف من عدة مئات من kilobites، ربما بعض الدقائق. وإن أهم معيار قياسي للضغط هو :

the JPEG (Joint Photographic Experts Group) File Interchange Format;

ومع هذا المعيار القياسي، يمكن تقليل ملف من 2mb إلى حوالي 100k، مع التضحية ببعض التفاصيل^(١٦). ولحسن الحظ فإن العين قادرة جدا على التحمل والخسائر ليست ملحوظة. ولنعت بعض الأرقام المخصصة، يعطى شكل Kodak Photo-CD حوالي مائة صورة من نوعية عالية الجودة على قرص مدمج؛ ويعطى نظام Protfolio ما يصل إلى ألف صورة ذات نوعية أقل جودة ولكن لازالت مقبولة، في حين أن الأساليب الفنية للضغط يمكن أن تعطي سبعة آلاف صورة على قرص مدمج واحد.

ويجب أيضا ألا نغفل المعايير القياسية للصناعة. وقد طورت : Adobe Sys- tems Inc، وهي التي طورت لغة التحكم في الطابع Postscript - طورت أيضا : Adobe Acrobat، وهو برنامج سوف يأخذ ملف طابع Postscript ويحوله إلى شكل يمكن قراءته على أى شاشة مرم GUI فى الشكل الأصلي، بشكل كامل مع الرسوم واللون^(١٧). وقد طورت Intel لوغاريتم Indeo للسيطرة على الفيديو

الرقمى وضغطه؛ وقد أدمج هذا فى فيديو ميكروسوفت الخاص بويندوز Micro-Quick Time for ، وفى Apple's Quick Time ، وفى Windows soft Video for Windows (١٨).

الانترنت

على مر السنين، أصبحت كميات كبيرة من المعلومات متاحة على الانترنت^(١٩)، وطورت برامج متنوعة لمساعدة المستخدمين لكى يشقوا طريقهم خلالها. ويجب أن نتذكر أنه لا توجد سيطرة عامة شاملة على الإنترنت، وأنه لا توجد هيئة مركزية تفرض النظام. وأى نظام يوجد هو نتيجة التعاون بين المستخدمين. وحينما بدأت الشبكة كان معظم المرور عبارة عن بريد الكترونى بين الأفراد، ولكن سرعان ما أصبح واضحاً أن جماعات قد بدأت تتشكل، تبادل المعلومات نفسها بين عدد من الناس. وقد أدى هذا إلى تكوين جماعات أخبار newsgroups؛ والبريد الذى يرسل ينتقل أوتوماتياً إلى كل الأعضاء. ويوجد الآن آلاف عدة من جماعات الأخبار حول العالم، كل منها له قائمة خدمة خاصة تدير آليات قائمة المشتركين، وعادة ما تتابع مسار ما يحدث لكى تحذف الرسائل «غير المناسبة» («غير المناسبة» قد تعنى ببساطة أنها خارج المجال - كل الرسائل تشغل حيزاً على القرص الصلب لوحدة تقديم الخدمة - ولكن بعض الرسائل قد يزعج المشتركين. الجدران ليست هى المكان الوحيد الذى يمكن للمرء أن يجد فيه graffiti!).

ولكى نستخدم الشبكة، فمن الضرورى أن تكون لدينا البرامج التى تتطابق مع المعيار القياسى TCP/IP:

(Transmission Control Protocol / Internet Protocol).

وتسمح تلنت login Telenet عن بعد إلى المواقع الأخرى للتعرف على المواد التى توجد هناك. ويمكننا FTP (File Transfer Protocol) من نقل الملفات

بين حاسبنا والحاسبت الأخرى. ولكى يشق المرء طريقه يمكننا استخدام برامج Gopher. وهناك عدة مئات من وحدات تقديم الخدمة تحتوى على قوائم هرمية (رتبية) تقود إلى المعلومات المتاحة؛ وهناك فهارس إتاحة عامة مباشرة لكثير من المكتبات يمكن الوصول إليها باستخدام تلتنت أو جوفر، منها المكتبة البريطانية ومكتبة الكونجرس^(٢٠). ومع برامج جوفر الخاصة بالعملاء يمكن خلق book marks لتحديد ذاتية المواقع التى قد يرغب المرء فى زيارتها مرة ثانية؛ وهذا يوفر قدرا كبيرا من كتابة الحروف! وهناك برامج أخرى مثل ARCHIE وفيروسنكا تستخدم كوسيلة لتحديد أماكن وحدات تقديم الخدمة أو الملفات.

ورغم أن TCP/IP هو البروتوكول القياسى من الناحية الواقعية لتبادل المعلومات، فإن هناك مشاكل التوافق مع ISDN، وكذلك مع المعيار القياسى الدولى الجديد لـ : أوسى (Open System Interconnection (OSI) ورغم أن آيم IBM أعلنت فى سنة ١٩٨٨ أنها سوف تبدأ تقديم منتجات بروتوكول أوسى OSI فى سنة ١٩٩٠، فإن أوسى OSI، وهو المعيار القياسى، لم يحل حتى الآن (١٩٩٥) محل المعيار القياسى القديم^(١٢)، وقد لا يحدث ذلك أبدا، فى ضوء الاستثمار فى TCP/IP.

شبكة العنكبوت العالمية

إن صعوبة إيجاد المعلومات على الإنترنت قد أدت إلى تطوير بروتوكول جديد لربط المواقع بالحاسب : HTTP : Hyper Text Transport Protocol .

وقد بدأ فكرة فى عام ١٩٨٩، ثم أدى إلى تنفيذ : شبكة العنكبوت العالمية World Wide Web فى : The European Centre for High Energy Physics CERN فى ١٩٩١. وقد عملت الشبكة بشكل كامل فى ١٩٩٣ مع تطوير برامج : Mosaic Web Browser بواسطة NCSA (National Center for Super-computer Applications) الذى وضعها كملكية عامة، بحيث يتسنى لأى أحد أن يحصل عليها باستعمال FTP^(٢٢). وقد كونت الوب ١٠٪ من المرور فى العمود

الفقرى NSF Internet فى مارس ١٩٩٣، بعد دخول موازيك؛ وفى سبتمبر ١٩٩٣ أصبحت تكون ١٪ وفى نوفمبر ١٩٩٤ أصبحت تكون ١٠٪؛ ومن الواضح أن الاستخدام مستمر فى النمو بسرعة كبيرة. ما الذى أدى إلى هذا الانفجار المفاجيء، فى استخدام الإنترنت؟

كان العامل الأول هو استخدام النص الفائق لبناء الروابط بين الوثائق. لقد كان الفيزيائيون فى CERN يجربون الحمل الزائد للمعلومات over-information load، واحتاجوا إلى طريقة أفضل لمتابعة مسار المطبوعات على الانترنت تلك التى وجدوا أنها مفيدة. وكان العامل الثانى استخدام الوسائط المتعددة؛ لقد كانت الانترنت تقتصر على النص، ولكن برامج الوب جعلت من الممكن استخدام الرسوم والتسهيلات التى يوفرها مرم GUI. والآن حل Netscape محل موازيك، وهو مستطلع رسوم graphics browser يمكن المستفيد من استخدام أساليب النقطة والطريقة (الطقطقة) click لكى يذهب إلى مواقع أخرى وكذلك لخلق روابط النص الفائق باستخدام bookmarks هناك برامج أخرى كذلك يوفرها موردون آخرون.) وكل موقع يتم إفراده بواسطة : a Univeral Resource Locator, URL، والذى يمكن أن يشتمل لاعلى مكان الحاسب فحسب ولكن على ممر الدليل إلى الملفات المخصصة.

وثمة عامل آخر ساهم فى نمو الاستخدام هو الاهتمام الذى أبداه البائعون التجاريون. ففى حين أن الإنترنت كانت إلى حد كبير مقصورة على تبادل النص بين المؤسسات الأكاديمية، فقد كان هناك اهتمام ضئيل من جانب هؤلاء الذين لا يريدون الوصول. ومع تطوير WWW، والتى أظهرت أنه يمكن استخدام الرسوم والصوت أيضا، فقد حدث اهتمام كبير جدا، من المدارس على سبيل المثال، ثم بدأت شركات متعددة فى تقديم الإتاحة على أساس دفع الأجرة. (رغم أن الاتصال على الإنترنت كان دائما مجانيا، أو «بتكلفة مكاملة محلية»، فقد تحملت التكلفة الجامعات، والحكومات، والهيئات التى تمولها الحكومة مثل NSF.) ومع تزايد الاستخدام، تزايدت التكاليف كذلك؛ وإن إتاحة الرسوم

(الصورة) والصوت تعنى أن قدرا أكبر من المعلومات يتم نقله الآن بين المواقع، مع تزايد الطلبات على تسهيلات الاتصالات عن بعد. وقد زادت الرابطة بين استراليا وأمريكا الشمالية فى سنة ١٩٩٥ إلى ضعف مقدرتها السابقة؛ وقد استغرقت الاستفادة الكاملة من المقدرة الإضافية ما يقارب اليوم!

وكما ذكرنا سابقا لاتوجد هيئة حاكمة على الانترنت، كما www. كذلك فلا توجد إلا سيطرة محدودة على المعلومات المتاحة عليها، كما على الطريقة التى تنظم بها المعلومات. وقد يحتاج الأمر إلى قدر كبير جدا من المهارة لتحديد أماكن كل مصادر المعلومات ذات القيمة فى مجال موضوعي معين، كما أوضح وسترمان Westerman^(٢٣)؛ وقد تعاون سبعة من المكتبيين العاملين فى مجال إدارة الأعمال لمدة طويلة لكى يفردوا مصادر المعلومات فى الإدارة على الإنترنت، ولكى يوفروا لمستفيديهم خدمة كانت صغيرة ومركزة بالقياس إلى الشبكة ككل. وقد اقترحت إمكانية استخدام نوت BSO (عالجنه فى الفصل ٢٠) للمساعدة فى تنظيم الشبكة؛ وقد ابتكر هذا التصنيف لإفراد المؤسسات عن طريق التغطية الموضوعية الشاملة، وربما أمكن استخدامه لعنونة مصادر المعلومات. ولانعرف حتى الآن إن كان مقدمو المعلومات على الإنترنت يريدون التعاون على هذا النحو. وربما أدى الحضور المتزايد للبائعين التجاريين إلى تحكم أوثق. والمستفيدون الذين يدفعون من أجل خدمة ما أقرب إلى أن يطلبوا سهولة الاستخدام من هؤلاء الذين يحصلون على الإتاحة مجانا!

المراجع

- McClure, C. R. *et al.*, 'Toward a virtual library: Internet and National Research and Education Network', *Bowker annual: library and booktrade annual*, 1993, 25-45.
- McClure, C. R. *et al.*, *The National Research and Education Network: research and policy perspectives*, Norwood, NJ, Ablex, 1991.
- MacColl, J. A., 'Library applications of a wide area network: promoting JANET to UK academic libraries'. *Information services and use*, 10 (3). 1990, 157-68.
- 2 Cawkell, A. E., 'The annual 'arrival' of speech recognition'. *Information services and use*, 10 (3), 1990, 133-4. (Editorial) Cawkell's scepticism is still justified, though progress is certainly being made.
 - 3 Auger, C. P., *Information sources in grey literature*, 3rd edn, London, Bowker-Saur, 1994.
 - 4 Yasui, H., *Desktop publishing: technology and design*, Chicago, Science Research Associates. 1989. This is one of the many books now available on DTP. Students should use a text which is conveniently available.
 - 5 *Infotrain* is an electronic journal produced by students of librarianship, available at <http://infotrain.magill.unisa.edu.au>
 - 6 Lancaster, F. W., *Toward paperless information systems*, New York, NY, Academic press, 1978.
 - 7 Bush, V., 'As we may think', *Atlantic monthly*, 176 (1), July 1945, 101-8.
 - 8 Nelson, T. H., *Computer lib: dream machines*, Redmond, WA. Tempus Books of Microsoft Press. 1987. This text is also available on the XANADU experimental machine.
 - 9 'Perspectives on the human-computer interface' [special issue], *Journal of the American Society for Information Science*, 43 (2), 1992, 153-201.
 - 10 American standard code for information exchange. American National Standards Institute X3.4: 1977.
 - 11 Honan, J. 'Highway more than a home shopping guide'. *The Australian*, June 20 1995. (Argues very strongly for the importance of text as opposed to graphics.)
 - 12 'Workshop on hypermedia and hypertext processing'. *Information services and use*, 13. 1993, 81-199. The need for standards is emphasized by G. Stephenson, 'Introduction', 85-7, and by M. Bryan, 'Standards for text and hypermedia processing', 93-102.
 - 13 Stern, D., 'SGML documents: a better system for communicating knowledge', *Special libraries*, 86 (2). Spring 1995, 117-24.
 - 14 Pfaffenberger, B., *World wide web bible*, New York, NY. MIS Press, 1995. Chapter 27: 'A quick introduction to HTML', 447-70.
 - 15 Popham, M., 'Use of SGML and HyTime in UK universities'. ref. 10 above, 103-9.
 - Burnard, L., 'Rolling your own with TEI', ref. 10 above, 141-54.
 - 16 Bryan, M. In ref. 12 above.

- 17 Fox, E. A. *et al.* 'Digital libraries', *Communications of the ACM*, **38** (4), April 1995, 23–8. (Introduction to a special issue on digital libraries, 23–109)
- 18 Pring, I. 'Video standards and the end user', *Information services and use*, **13**, 1993, 93–102.
- 19 Krol, E., *The whole Internet users' guide and catalog*, 2nd edn, Sebastopol, CA, O'Reilly and Associates, 1994. There are a number of good books on the Internet, but this is one of the best and most complete.
- Lynch, C. and Preston, C., 'Internet access to information resources' *Annual review of information science and technology*, **25**, 1990, 263–312
- 20 For the British Library, gopher portico.bl.uk. For the Library of Congress, telnet marvel.loc.gov. login as marvel. (Marvel is the LoC gopher.) To use the Library of Congress catalogue, telnet locis.loc.gov and follow the menus.
- 21 Cawkell, A. E., 'Videoconferencing, the Information Superhighway and the second Défi', *Information services and use*, **15** (2), 1995, 73–4. (In *Le défi Américain*, J. J. Servan-Schreiber argues the decline of Europe in the face of American Cultural imperialism.)
- 22 Books on the World Wide Web, of which ref 14 above is one example, are forming a publication explosion of their own. Many come with floppy disks or CD-ROM, containing software to enable users to set up their own home page. Not all home pages are of value.
- 23 Westerman, M., 'Business sources on the Net: a virtual library product', *Special libraries*, **85** (4), Fall 1994, 264–9.
- 24 CRG minutes, February 24 1995.

In order to keep pace with changes it is important to scan the computer section of a quality newspaper, and also read widely in the periodical literature; the main problem is not to become bogged down in trivia!

Appendix

Some of the relevant ISO standards are as follows:

ISO 7498:1988 Open systems interconnection reference model. (OSI)

ISO 8613:1989 Information processing – text and office systems – Office Document Architecture (ODA)

Part 1 Introduction and general principles

Part 2 Document structures

Part 4 Document profile

Part 5 Office Document Interchange Format (ODIF)

Part 6 Character content architecture

Part 7 Raster graphics content architectures

Part 8 Geometric graphics content architectures

Part 9 Audio Content Architecture

Part 10 Formal specifications

ISO 8879:1986 Standard Generalized Markup Language.

ISO 8879: 1988 SGML Supplement 1.

ISO 9069: 1988 SGML support facilities: SGML document interchange format (SDIF)

ISO 9541-1: 1991 Font information exchange: Part 1: Architecture.

ISO 9541-2: 1991 Font information exchange: Part 2: Interchange format.

ISO 9660:1987 Volume and file structure of CD-ROM.

ISO 10180: 1993 Standard Page Description Language (SPDL)

ISO 10744:1992 HyTime Hypermedia/Time-based structuring language.

ISO 10918 Joint Picture Experts Group (JPEG) – compression encoding for continuous tone pictures.

ISO 11172: 1993 Moving Picture Experts Group – digital moving picture compression method.

التكشيف المشتق ٢ : نظم الاتاحة إلى قواعد البيانات

خلفية

في الفصل ١ ، ألمحنا بإيجاز إلى أنه توجد ثنائية بين نظم الضبط البليوجرافي المستخدمة للكتب وتلك التي تستخدم للمواد الأخرى . وقد كانت نظم الكلمات الرابطة التي عالجتناها في الفصل ٣ تستخدم أصلا في فهارس المكتبات ، ولكن استخدام الحاسب جعل تطبيقها ممكنا على مواد أخرى كثيرة - مثال ذلك : مقالات الدوريات ، والتقارير الفنية ، وأعمال المؤتمرات . والتكشيف الإسنادى ، كما طوره جارفيلد ، كان يطبق بصورة مخصصة على مقالات الدوريات ، والكتب والمصادر الأخرى مثل براءات الاختراع قد تظهر في الكشاف الإسنادى باعتبارها مواد أو وحدات مسندة (أكثر الأعمال إسنادا هي الانجيل وشكسبير!) ، ولكنها لا ترد أبدا في كشاف المصادر . ونجد في الحقيقة أنه يوجد الآن نوعان مختلفان من قواعد البيانات البليوجرافية المحسبة ، تلك التي تعالج مايمكن أن يطلق عليه المطبوعات الكبرى - الكتب - وتلك التي تغطي المطبوعات الصغرى - وهى مقالات الدوريات وكل الأشكال الأخرى المشابهة من المطبوعات . وهناك فصل من الناحية العملية وهو أن الكتب يمكن أن تقف على رفوف المكتبة ، ولذلك يمكن ترتيبها فى نوع ما من الترتيب المفيد الذى هو عامل مهم فى البحث الموضوعى . أما المواد الأخرى فلا يمكن أن ترتب بهذه الطريقة . من الممكن أن ترتب الدوريات ككل على الرفوف ، أو أعمال المؤتمرات التى تظهر فى شكل الكتاب ، ولكن هذا لايعطى وصولا مباشرا إلى المقالات الفردية فى داخلها .

والعالم الحقيقي هو بطبيعة الحال رمادى اللون، لاهو بالأبيض ولا هو بالأسود، وهذا التقسيم الثنائي للأنظمة هو تبسيط، ولكن هذا الفصل العملى يوازيه الطريقة التى يعامل بها المجريان أو المساران لأغراض استرجاع المعلومات. فالكتب «تفهرس» بينما المواد الأخرى «تكشف». والأسلوبان يهدفان إلى الأهداف العامة نفسها: تحقيق ذاتية الوحدة (المادة) وتوفير الوصول إليها من خلال مداخل متعددة، بما فى ذلك الموضوع. ومع ذلك فإن فهرسة الكتب عادة تتضمن التلخيص summerization: فنحن نعالج المحتويات ككل، ونوفر الوصول إلى الموضوع على مستوى محدود - رقم تصنيف لترتيب الرفوف، ورأسا أو اثنين من رؤوس الموضوعات للوصول عن طريق الفهرس (الموضوعى). وينزع تكشيف المواد الأخرى إلى أن يكون أكثر تفصيلا: فليس لدينا رقم تصنيف لترتيب الرفوف، ولكن ربما للترتيب فى بليوجرافية مطبوعة، أو قد نكون أكثر كرما فى توفير المصطلحات للوصول إلى الموضوع. ولنعط مثلا عمليا:

The 5th International Study Conference on Classification Research

والذى أسندنا إليه كثيرا فى كتابنا هذا^(١)، هذا العمل يمكن فهرسته ككتاب، ويوضع على الرف فى ٠٢٥ر٤، ويعطى رأس الموضوع: الكتب - تصنيف. أما فى خدمة استخلاص مثل Library and Information Science Abstracts، أو فى مجلة تكشيف مثل Library Literature، فيمكن أن نتوقع مداخل لكل الفصول الفردية، عن تعد، تع، تمك، وبناء المكانز، وإعادة التصنيف وما إلى ذلك. وفى حين أننا قد نأسف على هذا الفصل، فإنه موجود منذ أن بدأت الكشافات والدوريات، وكانت له نتائج عملية بالنسبة لاسترجاع المعلومات. ونحن فى هذا الجزء ندرس قواعد البيانات التى تغطى المطبوعات الصغرى، مؤجلين دراسة فهرس الكتب - اليدوية والأوباك - إلى فصول لاحقة.

المطبوعات العامة

فى الفصل ٤، أشرنا إلى تطور جمع الحروف المحكوم بالحاسب فى أوائل

ستينات القرن ٢٠ وفي البداية كان هذا نادرا نسبيا، ولكن هناك ثلاث جماعات مهمة سرعان ما أدركت مزاياه. فهناك قواعد بيانات ببيوجرافية مثل Index med-icus الذى يمكن إنتاجه أسرع من الطرق التقليدية، ويمكن تركيبه بصورة مستمرة فى قواعد بيانات كبيرة يمكن بحثها ككل^(٢). وقد حل هذا مشكلة الضجر أو الملل الذى ينشأ من البحث فى عدد كبير من الإصدارات المطبوعة المستقلة. ولما كانت التجارب المؤقتة الأولى فى أواخر الستينات (ق ٢٠) قد أظهرت أن الوصول عبر الخط المباشر كان عمليا، فإن أعدادا متزايدة من قواعد البيانات الببيوجرافية قد أصبحت متاحة على الخط المباشر، وأصبح الوصول إلى ٤٥٠ منها ممكنا عبر خدمة ديايوج لاسترجاع المعلومات، وهى أكبر خدمة من نوعها، فى ١٩٩٤، وهذا العدد يعتبر صغيرا بالنسبة لجملة العدد المتاح الآن^(٣).

والجماعة الثانية التى استفادت من التحسيب هم ناشرو الصحف. يستطيع الصحفيون الآن إنتاج نسختهم على حاسب مصغر لكى تُدخَلَ مباشرة فى الحاسب الرئيسى، الذى يمكن استخدامه مباشرة لجمع الصحيفة كلها دونما حاجة إلى الخطوة الوسيطة وهى جمع الصحيفة على آلة لينوتيب. وفى الحقيقة فإن هذه الآلة، التى كانت أساس إنتاج الصحف فى القرن العشرين كله، قد اختفت من الاستعمال فى وقت يثير قصره الدهشة. وفى الأحوال العادية، تكون الصحف متاحة على الخط المباشر، والكثير منها يمكن الوصول إليه الآن بهذه الطريقة.

والجماعة الثالثة كانت الحكومات، وبدأت النظم الأساسية التشريعية وغيرها من الوثائق القانونية تنتج وتتاح على الخط المباشر بأعداد متزايدة لدرجة أنه يمكن القول أن ذلك قد أصبح الآن هو الشئ المعتاد، فى الدول المتقدمة. وقد كانت قواعد البيانات القانونية بين أوائل المصادر ذات النص الكامل التى نتاح على الخط المباشر^(٤). كذلك استخدم الناشرون التجاريون جمع الحروف المحكوم بالحاسب، ولكنهم لا يرغبون بطبيعة الحال فى أن يتيحوا مطبوعاتهم

مجانا! فعلى عكس الحكومات، يحتاج الناشر والتجارىون إلى أن يحققوا أرباحا من بيع منتوجاتهم. وسوف نتناول القضية المهمة : قضية الملكية الفكرية وحقوق الطبع فى الفصل ٢٨.

وفى الجملة، فإن هناك كميات كبيرة جدا من المعلومات على الخط المباشر يمكن الوصول إليها الآن عبر مصادر متنوعة، بعضها مجانا وبعضها على أساس تجارى. وحينما أصبحت قواعد البيانات على الخط المباشر متاحة لأول مرة حوالى ١٩٧٠، كان المستفيدون هم من المكتبيين وغيرهم من العاملين فى مجال المعلومات. وإن سهولة الوصول فى قواعد البيانات فى الوقت الحاضر، والبرامج سهلة الاستخدام (نسبياً) الموجودة التى يمكن استخدامها الآن، كل هذا يعنى أن أناسا كثيرين قد بدأوا يستخدمونها دون مساعدة من طرف ثالث. ومع ذلك فهناك مهارات معينة ينبغى معرفتها لتحقيق أفضل استفادة من المعلومات المتاحة لنا. وكما أشرنا فى الفصل ٢، فإن المستفيدين ليس لهم جميعا الطلبات نفسها للاستدعاء والصلة، ويمكن تعديل استراتيجيات البحث لتعرف ذلك. ويجب أن نكون على دراية أى قواعد البيانات يمكن أن تكون أكثر عائدا فى البحث عن أنواع معينة من المعلومات، وعلينا أن نبذل شيئا من الجهد لتقديم المعلومات التى يريدونها المستفيدون فى شكل يلبي احتياجاتهم.

وقد بدأ ديالوج، وهو أكبر خدمة على الخط المباشر فى ١٩٦٠، وقد استخدمه الوسيطان - المكتبيون وغيرهم من العاملين فى مجال المعلومات، والمستفيدون النهائيون - المستفيدون الذين يولدون أسئلة المعلومات. ومع ١٩٨٢، كانت أعداد الوسطاء والمستفيدين النهائيون الذين يبحثون فى هذه الخدمة متساوين تقريبا، ولكن مع ١٩٨٨ شكل المستفيدون النهائيون حوالى ٦٥٪ من المستفيدين الجدد^(٥). ومع ذلك فليس كل السائلين يرغبون فى القيام ببحوثهم بأنفسهم؛ وقد أظهرت دراسة فى جامعة كارولينا الشمالية أن عددا من المستفيدين طلبوا المساعدة، حتى رغم أن قواعد بيانات مختلفة كانت متاحة على قم CD (ERIC, Books in Print and Ulrich's plus) وكذلك BRS/After Dark (وهو

المعد خصيصا للمستفيدين النهائيين). ومن بين الأسباب التي أعطيت لتفضيل الأستعانة بوسيط، أن أكثر من الثلثين ذكروا الافتقار إلى الخبرة فى البحث وسجل ٤٠٪ أنهم لا يرغبون فى إضاعة وقتهم الخاص. والمستفيدون يمكنهم أيضا الاستفادة من الخيارين؛ وكثير ممن أجرى المسح عليهم خططوا لاستخدام وسيط مرة ثانية، ولكن يمكنهم أيضا أن يقوموا ببحوثهم الخاصة فى بعض المناسبات^(٦).

ولا يمكننا فى كتاب كهذا إلا أن نتناول الأساليب الفنية الأساسية فقط، ولكن كتبنا دراسة مثل كتاب ر. ج. هارتلى وآخرين. R. H. Hartley et al.^(٧). تدخل فى تفاصيل الأساليب الفنية للبحث بطريقة ليست ممكنة بالنسبة لنا هنا، حيث يمكننا أن نحاول فقط أن نلخص الأساليب الفنية الرئيسية التى من المؤكد أنها تستخدم. ولما كانت قواعد البيانات متاحة على الخط المباشر عبر خدمات مثل ديالوج، أو BRS/Search أو ISA/IRS، ولكنها يمكن أن تكون أيضا على قم CD أو عبر الإنترنت فى الوقت نفسه مع Dalrymple و Roderer^(٨) - لكل هذا فقد استخدمنا العنوان : الإتاحة (الوصول) إلى قواعد البيانات وليس البحث على الخط المباشر.

بحث الحاسب للنصوص

إذا عرفنا أن بعض المعلومات التى نريدها توجد فى وثيقة معينة، ولكن لاتوجد قائمة محتويات أو كشاف، فلكى نصل إلى المعلومات علينا أن نقرأ الوثيقة كلها، باحثين عن الكلمات التى تهمننا. وسوف نستخدم المصطلح الذى استخدمناه سابقا : نحن نحاول أن نضاهى match حاجاتنا كما تعبر عنها كلمات معينة، مع الكلمات التى نريد الوصول إليها فى الوثيقة. قد نكون مستعدين لأن نفعل هذا بالنسبة لوثيقة واحدة، ولكن حينما نبدأ التفكير فى الأمر على ضوء مجموعة من الوثائق فإن من الواضح أن العملية تصبح غير ممكنة من الناحية العملية. وحتى لو بحثنا فى قائمة طويلة من العناوين فسوف يكون هذا مضیعة

للوقت، وهو أمر اكتشفه قراء كثيرون من تصفحهم لقوائم بائعى الكتب والمطبوعات المشابهة.

ويمكن للحاسب أن ينفذ هذا النوع من المضاهاة بسرعة عالية؛ فإذا كانت العناوين أو أجزاء النص فى شكل مقروء للآلة، كما هو معتاد الآن، فإننا يمكن أن نبرمج الحاسب لكى ينفذ عملية المضاهاة ويحقق ذاتية الوثائق التى من المنتظر أن تكون مفيدة لنا. وكل ما ينبغى علينا عمله هو أن نغذى الحاسب بالكلمات التى نريد مضاهاتها. ومع الإتاحة إلى كميات كبيرة من النصوص سواء كانت على الخط المباشر أو على قم، فإن الوصول إلى إجابة عن سؤال معين يُصبح مسألة استخراج سريع وغير مؤلم. فالحاسب لا يتعب من البحث - فليس عنده نقطة اللاجدوى! - كما أنه لا يضجر أو يسأم. وهو ببساطة يستمر فى البحث عن المعلومات التى قلنا إننا نريدها حتى يستكمل البحث. (يمكننا بطبيعة الحال أن نكتشف عند هذه المرحلة أن ماقلنا إننا نريده لم يكن فى الحقيقة صياغة دقيقة لاحتياجاتنا! وغالبا ما تتضمن صياغات البحوث أخطاء، فى حين أن الكثير منها لا يمكن تتبعها للحصول على «أفضل» نتيجة).

كانت نظم استرجاع المعلومات الأولى تعمل بنظام الدفعات batch، واستخدمت الشريط الممغنط كواسطة تخزين. وكان هذا يعنى أن البحث يسير ببطء؛ وحتى قاعدة بيانات صغيرة تتألف من عدة آلاف من المواد كانت تحتاج لبحثها إلى عدة دقائق، حيث كان البحث يتم بطريقة التتابع (قارن سرعة تحديد مكان مسار على LP أو قم CD بسرعة البحث فى شريط كاسيت!) وقد أدى دخول تخزين الوصول العشوائى random access storage إلى تحسين كبير جدا فى وقت الوصول، ولكنه ظل يحتاج إلى بحث كثير لكى يصل إلى الوحدات التى تضاهاى الطلب. وكان الحل هو خلق ملف ثان مقلوب أو معكوس inverted يسجل فيه كل مصطلح فى الأصل (مع استثناء الألفاظ المسجلة فى قائمة التوقف) يسجل كمدخل كشاف. وحينئذ يبدأ البحث فى ملف الكشاف، الذى يعطى العدد والأماكن التى رصد فيها المصطلح، ويمكن بسرعة أن يقارن

الأماكن المرصودة تحت مصطلحين اثنين لتحديد أى المرصودات أكثر اتفاقاً مع الاثنين. فى بعض النظم، يسجل ملف مقلوب واحد المصطلحات التى لها عدد من نقط الرصد postings، ويسجل ملف آخر الأماكن المخصصة. وتستخدم الطريقة نفسها أساساً مع الملفات الأكبر لأيامنا هذه. وإن ثمن تحسين أوقات الوصول هو الحاجة إلى ملفات كشافات إضافية، وهذا يعطى قاعدة بيانات ربما يصل حجمها إلى ضعف الأصل. وتحتاج خدمات مثل DIALOG إلى فئات من gigabytes من التخزين لكى تستوعب مجموعاتها؛ وتشغل ملفات الكشف المصاحبة لقواعد البيانات جزءاً كبيراً من طاقة التخزين. وتستفيد أغلبية قواعد البيانات الحالية من طريقة الملف المعكوس، ولكن التحليل التجميى cluster analysis والعمليات المشابهة تنفذ على النص كما هى، دونما حاجة إلى الملفات المعكوسة.

استراتيجيات البحث

يمكننا أن نبرمج الحاسب لكى ينفذ برنامج المضاهاة بطريقتين، توازيان التغييرات التى طرأت فى استخدام البرامج الأخرى للحاسب مثل معالجة الكلمات. والبرامج الأولى لقواعد البيانات احتاجت من المستفيد أن يكتب الأوامر Commands بلغة خفيفة esoteric بشكل يزيد أو ينقص، مثال ذلك : str بدل stringsearch. وهذا يعنى تعلم الأوامر قبل أن يكون الشخص قادراً على الاستفادة جيداً من الخدمة. ولما كانت كل خدمة تستخدم لغة أوامر مختلفة، فإن هذا كان يمكن أن يؤدى إلى اللبس؛ ولحسن الحظ، فإنه لما كانت لغات الأوامر تؤدى الواجبات نفسها، فإن المرء حينما يتعلم لغة واحدة لا يكون من الصعب عليه أن يسيطر على الأخرى. ولقد جرت محاولات لتطوير «لغة أمر عامة»، بحيث تستخدم كل الخدمات الأوامر نفسها، ولكن لم يكتب النجاح لهذه المحاولات⁽⁹⁾. والبديل لاستخدام الأوامر هو استخدام القوائم menus؛ وهذا أبسط فى الاستعمال، ولا تحتاج إلى تعلمها، ولكنه أبطأ من استخدام

الأوامر مباشرة. (نحن نتحدث فقط عن ثوان، ولكن المستفيد على الخط المباشر يفضل أن يملأ الدقيقة بما يساوي ستين ثانية من التشغيل! والمبتدئ قد يرغب هو أيضا في أن يستفيد من الشاشات المساعدة Help، وهذا يبطل من العملية، إذا لم تكن القوائم مشروحة بذاتها بشكل كامل.) وتوفر خدمة ديا لوج وكذا الخدمات الأخرى الاختيار لعدد من قواعد البيانات؛ وحينما يألف المستفيدون النظام من خلال استخدام القوائم، فإنهم يمكن أن يتحولوا إلى استخدام الأوامر.

كان يفترض حينئذ أن يصبح البحث على الخط المباشر بسيطا جدا؛ ولكنه من الناحية العملية يصبح أكثر تعقيدا^(١٠). لنفرض أنا مهتمون بموضوع التصنيف Classification؛ قد نجد الوثائق التي تستخدم اللفظ Classifying مفيدة بالدرجة نفسها. ومرة أخرى لنفرض أن اهتمامنا هو بالتلوث pollution - وهو موضوع يدور كثيرا في العقل العام. يمكننا أن نبحث عن pollution، ولكننا حينئذ نكون قد أخطأنا الوثائق التي تستخدم فقط كلمة pollutant، أو polluting والحل بسيط : يمكننا أن نبتز truncate المصطلح الذي نبحث عنه ليصبح pollut. وقد نحتاج إلى أن نبين بصفة مخصصة أننا قد استخدمنا البتر وذلك بإضافة رمز مثل النجمة*، لكي نبين أننا مهتمون بالساق plus، وإلا فإننا قد نجد أننا قد حصلنا على الاجابة صفر nil؛ وبعض النظم تأخذ أوتوماتيا أى مدخلات على أنها بداية مصطلح البحث، ولا تتطلب دالة مخصصة. ويمكننا أن نستخدم البتر إلى الأمام (الأمامي)، مثال ذلك *POLLUT، أو إلى الخلف (الخلفي)، مثال ذلك *CLASSIF*؛ والأخير سوف يأتي أيضا باللفظ : reclassification بالإضافة إلى الألفاظ التي تبدأ بـ classif... ولكنه سوف تأتي أيضا بـ declassification، وهو عملية جعل الوثائق السرية عامة. وإن البحث عن معلومات عن دور الأبوين يمكن أن يصاغ على أنه *PARENT*، وهذا يسترجع :

parents, parental, parenthood, grandparent, grandparents

ولكنه سوف يسترجع كذلك : parenthesis, parenteral, transparent!

فى بعض النظم سوف يعرض الحاسب قائمة بالمصطحات التى سوف يأتى بها البتر، والتى قد تقود إلى أفكار ثانية! ويجب أن نكون قادرين أيضا على استخدام a wildcard لكى نسمح بالاختلافات فى الهجاء، مثال ذلك : WOM?N سوف يسترجع Woman, Women ؛ F?ETUS سوف يسترجع fetus (الهجاء الأمريكى) أو foetus (الهجاء البريطانى).

ومن الممكن أن يكونا لدينا برنامج تحليل لغوى parsing program يتعرف اللواحق suffixes ويقوم بعملية البتر أوتوماتيا. ويعرف هذا باسم التجذير stem-ming، ومن المهم أن نعرف ما إذا كان البرنامج الذى نستخدمه سوف يفعل هذا، أو أنه سوف يتوجب علينا أن نخصص دالة البتر مثل* التى بينها سابقا. فى الحالة الأولى، سوف يكون POLLUT مقبولا كمصطلح بحث، ولكن فى الحالة الثانية سوف يكون مرفوضا إذا لم يشتمل على النجمة. ويمكن أن يحدث التجذير فى مرحلتين؛ الأولى على مستوى بسيط، وعلى سبيل المثال يمكن أن يحول الجمع إلى مفرد، ويخذف اللواحق التى تمثل الأشكال الفعلية (-ing,-ed) ويدمج الهجاءات المختلفة، (-isation,- ization;- our-or). والمرحلة الثانية وهى الأقوى قد تحذف قدرا كبيرا من اللواحق، مثال ذلك : -itis,-able ويمكن أن يكون التجذير مساعدا على البحث ذات قيمة، ولكن كما هو الحال فى البتر اليدوى، فإنه يمكن أن يقود أيضا إلى مصيبة!^(١١).

لقد افترضنا حتى الآن أن البحث يتألف من لفظ واحد، ولكن من الناحية العملية سوف نفكر فى العادة فى أكثر من لفظ واحد يدل على الموضوع الذى نهتم به. مثال ذلك : قد نهتم بالموضوع تلوث الماء water pollution وليس التلوث ككل. فى هذا الوضع يمكننا الاستفادة من مقدرة الحاسب على تداول الصيغ المنطقية، والمنطق المستخدم هنا هو المنطق البوولى Boolean وليس الأرسطى، وهو يعنى أننا يمكن أن نربط الكلمات التى نبحث عنها بواسطة المعاملات operators و، أو، وليس AND, OR and NOT. ويصبح بحثنا:

WATER AND POLLUT*

However, we might remember that water includes sea(s) and river(s), and modify the search to take account of this:

POLLUT* AND (WATER OR SEA* OR RIVER*)

We might want to exclude sewage as a form of pollution:

POLLUT* AND (WATER OR SEA* OR RIVER*) NOT SEWAGE

The order of precedence among the operators is NOT, AND, OR. Thus

POLLUT* AND WATER OR SEA*

ومع ذلك، فقد نتذكر أن الماء يشتمل على البحر (البحار) والنهر والأنهار وهكذا نعدل بحثنا بحيث يضع ذلك في الحسبان :

POLLUT* AND (WATER OR SEA* OR RIVER)

وقد نرغب في أن نستبعد الصرف SEWAGE كشكل من أشكال التلوث :

POLLUT* AND (WATER OR SEA* OR RIVER*) NOT SEWAGE

وإن ترتيب الأسبقية بين المعاملات هو : ليس، و، أو. وهكذا فإن :

POLLUT* AND WATER OR SEA*

سوف تعامل على أنها بحث عن pollut* مرتبطا بالماء، أو البحر sea* مستقلا.

وإن استعمال الأقواس في المثال السابق ضروري لتجنب ذلك، وهو ممارسة جيدة يوضح صياغة البحث على أية حال. ويجب أن نتذكر كذلك أن المستفيدين غير المتعادين على الأساليب البولية قد يستخدمون المعامل الخطأ؛ فإذا كانوا يريدون معلومات عن Cats and dogs مثلا، فإنهم يحتاجون إلى أن يخصصوا هذا على أنه "cats OR dogs" وإلا فإنهم سوف يسترجعون فقط المعلومات التي تعالج الاثنين، وليس المعلومات التي تعالج كلا على حدة.

ويمكن أن نصوغ المثال السابق في شكل سلسلة من البحوث البسيطة، ونستخدم النتائج عند كل مرحلة كمدخلات في المراحل التالية، كما يتضح هنا (استجابة النظام تحتها خط):

1 POLLUT*
SEARCH 1 POSTINGS 732
2 WATER
SEARCH 2 POSTINGS 1653
3 SEA*
SEARCH 3 POSTINGS 451
4 RIVER*
SEARCH 4 POSTINGS 679
5 2 OR 3 OR 4
SEARCH 5 POSTINGS 2215
6 1 AND 5
SEARCH 5 POSTINGS 142
7 SEWAGE
SEARCH 7 POSTINGS 284
8 6 NOT 7
SEARCH 8 POSTINGS 114

وتطبق المعاملات AND، OR، و NOT على كل الوثائق؛ في المثال السابق قد نبحث عن وثائق يأتي التلوث فيها مع الماء أو البحر أو النهر، ولكن لم يرد الصرف. وقد نرغب في أن نكون أكثر تحديدا، ونخصص أن الكلمات تظهر في الفقرة نفسها، والجملتها نفسها، أو المجاور أو القريب adjacent نفسه، وبعض الخدمات تسمح بهذا. لاحظ أن ADJ قد يعدل أو يحدد، مثال ذلك (n) ADJ، حيث n هي الفصل الذي نريد أن نقبله. مثال ذلك: (5) ADJ تعني أن المصطلحين المبحوث عنهما يجب أن يردا في نطاق خمس كلمات مع بعضهما.

WATER AND POLLUT*
WATER SAME POLLUT*
WATER WITH POLLUT*
WATER ADJ POLLUT*

(same document)
(same paragraph)
(same sentence)
(adjacent)

وقد يكون ترتيب الكلمات مهما أيضا؛ فلو أننا خصصنا :

"retrieval of informa- INFORMATION ADJ (2) RETRIEVI*
tion" لأن الكلمات ليست على الترتيب المخصص، أو "information we want
to retrieve"، لأن الكلمات سوف تكون متباعدة. هل يأخذ البرنامج الذي
نستخدم في الحساب كلمات التوقف عند تحديد مدى قرب الكلمات؟ قد نحتاج
إلى أن نفكر مليا حول ترتيب الكلمات، وإمكانية أن تفصل عن بعضها
بواسطة كلمات غير مبحوث عنها، عند صياغة بحثنا. وهناك بديل ممكن هو أن
توضع عبارات البحث بين علامتى تنصيص «...» حتى يتسنى معاملتها ككل.
قد نبحث عن مواد عن "circulation" AND "control"، لكي نجد أننا قد
استرجعنا:

Control of the peripheral circulation in man!

ويمكن أن يكون البحث المتجاور proximity مفيدا أيضا إذا لم نكن متأكدين
هل نحن نبحث عن كلمة أو اثنتين، مثال ذلك: end user - coordi- أو post-
nate^(١٢). ولكي يكون البحث المتجاور ممكنا، فمن الضروري أن يسجل
الملف المقلوب مكان المصطلحات بشكل دقيق جدا، ولذلك فإنه يصبح أكبر
من ملف يسجل ببساطة أن المصطلح يرد في مكان ما somewhere في الوثيقة.

وقد نخطئ في تهجى كلمة نستخدمها كمصطلح بحث، إما بسبب الجهل أو
بسبب الافتقار إلى مهارات الكتابة (على لوحة المفاتيح)، أو قد نستخدم كلمة
غير مستعملة؛ في هذه الحالة فإن النظام يعرض قائمة بالكلمات القريبة من
هجائنا في ترتيب ألفبائي، أو يصدر الصوت نفسه (الوزن أو السجع -asso-
nance)، وهذا يمكننا من تصحيح الخطأ، أو اختيار مصطلح آخر. (هذا مشابه
جدا لعملية مراجعة التهجى في معالج الكلمات).

وهناك طريقة قوية للتصحيح الأوتوماتى للأخطاء وهو التصحيح الذى
يستخدم كودات Soundex؛ وهذا يتضمن حذف كل ماعدا الحروف اللينة الأولى
وضغط النتيجة إلى أربعة حروف. وهى تتضمن معجما كبيرا وقدرًا كبيرا من

التشغيل، ولكنها يمكن أن تتعامل مع معظم الأخطاء سوى تبديل الوضع trans-position - والذي هو خطأ شائع بالنسبة للكثيرين من الكتبة غير المهرة^(١٣).

وقد تتضمن قاعدة البيانات التكشيف بواسطة لغة مقيدة وكذلك كلمات النص، أى: مصطلحات يضيفها المكشف من قائمة سبق تحديدها. فإذا كانت هذه هي الحالة، فينبغى أن نكون قادرين على عرض مصطلحات البحث من اللغة، وكذلك أية مصطلحات مترابطة related، أو نوسع بحثنا لكي يشمل عليها؛ وهذا يتضمن التوجيه المزعج فَجْرُ EXPLODE فى مدلاين! وبصفة عامة، يعرف هذا الأسلوب بأنه استخدام الحواجز أو الحدود hedges؛ يمكن جمع مجموعة من المصطلحات معا «فى نطاق حاجز hedge» لتمثيل موضوع أوسع لا يوجد له مصطلح مفرد مناسب، وذلك عن طريق استخدام المعامل المنطقي OR. مثال ذلك:

USA OR France OR China OR Russia OR United Kingdom

يمكن أن تمثل «الأعضاء الدائمين فى مجلس الأمن التابع للأمم المتحدة». وتمثل الحواجز أو الحدود نوع التجميع الذى نجده فى خطط التصنيف أو اللغات المقيدة المشابهة^(١٤)، واستخدامها كما هى أثناء بحث النص قد يسبب مشكلات. قد يكون لمصطلح البحث مرادفات أو أشباه مرادفات، مثال ذلك:

drunk driving, drink driving, driving under the influence, drunken driving...;

أو يكون متعدد المعانى ويحتاج إلى سياق يوضحه، مثل : records؛ أو قد يكون ردىء التعريف، مثل : democracy، أو قد يرد بكثرة إلى حد أنه لا يعطى نتائج بحث مفيدة^(١٥). ويمكن أن تخلق الحواجز بواسطة منتجى قواعد البيانات، كما فى مدلاين، أو بواسطة الباحثين، الذين قد يعدلونها على ضوء التجربة. وقد تبني على العلاقات الدلالية الموجودة فى مكتز أو معجم، أو على التواتر الذى يبينه تشغيل الحاسب.

وتمثل الوثائق فى الحاسب بواسطة تسجيلات records؛ ومجموع التسجيلات من النوع نفسه تصبح قاعدة بيانات database؛ وتقسم تسجيلات قاعدة البيانات عادة إلى حقول fields، وهذا يمكننا من أن نقصر بحثنا على مجال واحد بالذات. مثال ذلك : إذا كنا نعرف اسم المؤلف يمكننا أن نختصر وقت البحث عن طريق بحث حقل المؤلف فقط، أو نبحث حقل العنوان إذا كنا نعرف العنوان كله أو جزءا منه. وإذا كانت قاعدة البيانات تشتمل على واصفات من لغة مقيدة فى حقل الواصفات فإننا يمكن أن نستفيد من هذه بطريقتين.

يمكننا أن نختار الواصفات من اللغة المقيدة لكى نجرى بحثنا؛ أو يمكننا بدلا من ذلك، إذا ما وجدنا وثيقة مفيدة، مثلا من خلال البحث عن كلمات فى العنوان، يمكننا أن نستخدم الواصفات المستخدمة لتكشيفها لكى نراجع استراتيجيتنا للبحث. وبعض الوثائق قد يكون لها حقل للمستخلص، أو حقل لنصها الكامل؛ وإن بحث النص الكامل للوثائق على الخط المباشر قد يستغرق وقتا ويضيف إلى الأجرة - وهى حجج جيدة لاستخدام قواعد البيانات على أقراص مدمجة!

وإذا كنا ننفذ بحثا يحتوى على كلمات متعددة، فمن الممارسات الجيدة أن نعالج كل كلمة على أنها بحث مستقل، لكى نحصل على نقاط رصد لكل منها، كما أوضحنا سابقا. وحيث أن نختصر وقت البحث بطريقة فعالة للغاية عن طريق البدء بالمصطلح الذى له أقل عدد من نقاط الرصد ونوفقه مع التالى له فى العدد (الأقل). ويمكننا حتى أن نضغط عدد الاصابات hits إلى عدد مقبول دون أن نبحث عن المصطلح الأكثر استخداما، وهو الذى من المرجح أن يكون الأقل فاعلية فى رفض الإحالات غير المرغوبة.

مشكلات البحث البوولى

كل الذين يستخدمون البحث البوولى سرعان ما يعتادون على واحد من عيوبه. ولكى يكون البحث ناجحا فلا بد أن تقع النتيجة فى نطاق نقطة العتب أو

اللاجدوى $futility\ FP$ ؛ بالنسبة للمستفيد فإذا كانت FP هي m ، والبحث يسترجع n من الوثائق، فإنه سوف يكون ناجحا فقط إذا كانت $n \leq m$ ^(١٦). وإن البحث في قاعدة بيانات كبيرة عن مصطلح واحد كثيرا ما يعطى عددا من نقاط الرصد أكبر بكثير من اللازم، وهكذا نستخدم AND لمصطلح ثان؛ فإذا أعطى أكثر من اللازم فإننا نستخدم AND مع ثالث، وهكذا. ولسواء الحظ، فإننا غالبا مانجد أن إضافة مصطلح آخر أكثر إلى صياغة البحث يقلل عدد الإصابات من عدد كبير لدرجة غير مقبولة إلى صفر. وبدلا من ذلك، يمكن أن نبدأ بحثنا عن طريق تخصيص عدد من المصطلحات لكي نستخدم كلها مع AND ، لكي نجد أننا لانسترجع شيئا على الإطلاق. كيف يمكن أن نعيد تنظيم بحثنا بشكل أمثل لكي يعطى نتيجة مقبولة؟ إن الطريقة في الوضعين معا هي أن نسقط مصطلحا أو أكثر، ولكن أى المصطلحات؟ الاتجاه هو أن نحذف المصطلحات التي يعتقد الباحث أنها الأقل من حيث الأهمية، ونستبقى تلك التي تعتبر الأكثر من حيث الأهمية.

وتوصف هذه الطريقة بأنها «تأثير المثبت $anchoreffect$ »، وهي في الحقيقة تستبقى مصطلحات معينة في كل متغير من الصياغة الأصلية حينما يؤدي حذفه إلى النجاح.

وثمة طريقة لتداول المصطلحات المختارة هي البحث عن توافيقها $Combinations$ ^(١٧). بالنسبة لـ n من المصطلحات هناك $2^n - 1$ من التوافق، وعلى هذا فإننا إذا بدأنا بخمسة مصطلحات فسوف يكون هناك ٣١ توفيقا ممكنا. وكل ثلاثة مصطلحات سوف تعطينا سبعة توافيق. ولكي نعمل كل هذه بطريقة عقلية ونجرب كل واحد فسوف يستهلك هذا وقتا طويلا كما أنه عمل مرهق، ولكن الحاسب يمكنه أن يعمل ذلك لنا، مستخدمين نوع استراتيجية البحث الفعالة نفسها التي ذكرناها آنفا، وهو يعطينا النتيجة عند كل مرحلة. وعلى هذا فبوسعنا أن نوقف البحث في حين تكون النتيجة لازالت تحت نقطة اللاجدوى، بينما نظل واثقين من أن نتائج البحث سوف تتوافق مع بحثنا إلى درجة قريبة إن لم تتوافق تماما. وبعض نظم استرجاع المعلومات تقدم هذه الفائدة، وتعرف باسم

بحث النخبة أو المجموعة المختارة quorum searching، رغم أنها هي وحدها أكثر البرامج المتطورة التي تملك قدرة التحسب اللازمة.

المخرجات المرتبة أو الموزونة

إن البحث البولي لايعطينا أى سيطرة على نقطة إيقاف البحث cutoff point . وبمعنى آخر، لايمكننا أن نقول كم عدد الوثائق التي نحب أن نسترجعها ونعمل لتحقيق ذلك المستهدف؛ وعلينا أن نقبل مايعطيه لنا الحاسب. وثمة معالجة أكثر فاعلية وهي أن نستخدم إجراءات بحثية أكثر تطوراً، بحيث يمكن أن نرتب نتائج بحث ما حسب درجة الصلة المحتملة. وحينئذ يمكننا أن نختار العدد الذي نريده من قمة القائمة وهكذا يكون لدينا شيء من الثقة فى أن هذه التي اخترناها هي الأكثر فائدة. وبالمقارنة مع البحث البولي، حيث يملئ النظام نقطة إيقاف البحث، فإن هذا يمكننا من تحديد هذه النقطة: يمكننا أن نقول إننا نريد أن نرى الستة، أو المائة، أو الثلاثة وثائق الأكثر عطاء، ونتجاهل الأخريات إلا إذا وجدنا أننا نحتاج إلى توسيع البحث. وبهذا لا نكون تحت رحمة النظام ولكن يمكن أن نضع مقاييسنا (بارامترات) للنجاح.

وإن إحدى الطرق لتحقيق هذا هو أن نزن weight المصطلحات المستخدمة للتكشيف أو البحث، ولكن كيف نوزع الأوزان المناسبة؟ يمكننا إما أن نفعل ذلك بطريقة عقلية أو من خلال الحاسب. وفى كلتا الحالتين يمكننا أن ننظر إلى رابطة مصطلح الوثيقة document - link على أنها مصفوفة matrix يكون فيها x من الوثائق كشف بواسطة y من المصطلحات؛ w_{ij} هو المدى الذى استخدم به المصطلح i التكشيف الوثيقة j.

| | D o c u m e n t s | | |
|---|-------------------|-----------------|-----------------|
| T | w ₁₁ | w ₁₂ | w ₁₃ |
| e | w ₂₁ | | |
| r | w ₃₁ | | |
| m | | | |
| s | | | |
| | | | w _{ij} |

وفى البحث عن المصطلحات غير الموزونة، فإن القيمة فى كل خلية إما أن تكون 1 (ذلك المصطلح استخدم لتكشيف تلك الوثيقة) أو صفر (ذلك المصطلح لم يستخدم لتكشيف تلك الوثيقة). هذا هو الوضع الذى نجده فى البحث البولى.

فى نظام البحث الذى يستخدم الأوزان، يمكن أن تتراوح قيمة wij من صفر إلى ١، ويمكن اعتبارها احتمالية أن مصطلحا بعينه سوف يكون مفيدا فى استرجاع وثيقة ما إجابة لسؤال معين^(١٨). ويمكن توزيع الأوزان عقليا إذا كانت المجموعة صغيرة ولكن سرعان ما يصبح ذلك غير عملى. ولذلك فإن علينا أن ندرس الطرق التى يمكن بواسطتها أن توزع الأوزان بواسطة الحاسب.

إحدى الطرق هى أن نستخدم الطرق الاحصائية للدلالة على أهمية المصطلحات. وهذا يمكن عمله بطرق مختلفة. وإن عدد مرات ورود اللفظ هو بحد ذاته مبسط إلى حد ما، ولكنه يكون أقوى إذا أخذنا ورود اللفظ بالنسبة إلى الورد المتوقع، اعتمادا على أعداد الألفاظ فى جسم كبير من الإنتاج الفكرى. وقد طورت طريقة أخرى، هى (the Assoiative Interactive Dictionary (AID، لبحث الملفات المختلفة لمدلارز^(١٩). ويبين لنا الملف المقلوب لقاعدة البيانات المصطلحات المستخدمة وعدد نقاط الرصد لكل منها؛ ومن هذه يمكن حساب الورد المتوقع فى أى مجموعة معطاة من الوثائق الخاصة بمصطلح معين. دعنا نفترض أن بحثا ما استدعى n من الوثائق من الإجمالى N فى المجموعة. بالنسبة لأى مصطلح معين يردن فى هذه الوثائق يمكننا أن نجد العدد الاجمالى لنقط الرصد لذلك المصطلح فى المجموعة T. وعلى هذا يكون الورد المتوقع E هو :

$$E = T.n/N$$

فإذا حسبنا الآن العدد الحقيقى لمرات الورد صفر، فإننا يمكن أن نصوغ مقياسا للغلاقة (x^2 the statistical measure) cf لكى نبين قوة الارتباط بين

المصطلح المدروس والوثيقة المسترجعة:

$$R = (O - E) / E$$

ويمكننا عندئذ أن نحسب قيمة R لكل واحد من المصطلحات التي ترد في مجموعة الوثائق، ونضعها حسب رتبها. قد يدخل مستفيد ما مصطلح بحث، ويطلب رؤية المصطلحات المترابطة؛ والنتائج مشابهة لتلك التي وجدت من خلال الجهد العقلي، ولكنها تشمل على بعض النتائج غير المتوقعة. ومن الأمثلة واحد بدأ بالكلمة "Shell fish" وحدد مكان عدد كبير من المصطلحات المرتبطة، وكانت المصطلحات التسعة الأعلى في الرتبة هي:

| Rank | XTRA-PSTGS | Term |
|------|------------|------------------|
| 1 | 390 | OYSTERS |
| 2 | 334 | MUSSELS |
| 3 | 227 | CLAMS |
| 4 | 185 | TIDES |
| 5 | 180 | ESTUARIES |
| 6 | 143 | PARAHAEMOLYTICUS |
| 7 | 138 | CRASSOSTREA |
| 8 | 101 | SEAFOODS |
| 9 | 88 | VIRGINICA |

وفي حين أن معظم هذه قد يكون متوقعا، فإن بعضها لم يظهر في قائمة مشتقة بالجهد العقلي! ومع ذلك، فمن المؤكد أنها مبررة وذلك عن طريق ورودها في الوثائق المدروسة؛ وكما أشار (ت) سفينونيس Svenonius، فإن الارتباطات المشتقة من تحليل النص الكامل بواسطة الحاسب يمثل النهاية المنطقية للسند الأدبي^(٢٠)!

وفي حين أن ورود الكلمة المطلقة ليس بالقوة نفسها لطريقة الورود المتوقع للفظ، فلا يزال له قيمة. ومن الواضح أن الألفاظ التي ترد كثيرا جدا في وثيقة

مجموعة معينة سوف تستدعي عددا كبيرا من الوثائق، فإن الكثير منها سوف تكون متصلة اتصالا هامشيا بسؤال معين، أو لا تكون متصلة على الإطلاق؛ والألفاظ التي لا ترد كثيرا سوف تعطى استدعاء منخفضا، وبهذا تمكنا من رفض الوثائق غير المسنودة بصورة أسهل. وعلى هذا فإن عامل الوزن يمكن أن يبنى على العلاقة العكسية لكثرة الورود (التواتر).

ويمكن حساب عامل تمييز اللفظ term discrimination factor لتحديد أى المصطلحات سوف تكون الأكثر فائدة فى تمييز وثيقة عن غيرها. ولو أن عندنا وثيقتين يمثل كلا منهما مجموعة من مصطلحات الكشف، فإننا يمكن أن نحسب مقياسا للتشابه بينهما؛ فحيثما تكون مجموعات مصطلحات الكشف هى نفسها، فإن مقياس التشابه سوف يكون ١، أما إذا لم يكن ثمة شىء مشترك بين المجموعتين، فسوف يكون المقياس صفرا. ويقع المقياس عادة فى مكان ما بين الصفر و ١. ويمكن أن نحسب مقياس التشابه لمجموعة من الوثائق لكى نصل إلى رقم المتوسط. ويمكن حينئذ إعادة حساب الرقم مع كل مصطلح ثم حذفه بدوره؛ وسوف يتبين عامل التمييز لكل مصطلح عن طريق الفرق بين مقياس التشابه الذى حصلنا عليه مع ذلك المصطلح الذى حذف ومتوسط مقياس التشابه، ثم ترتب المصطلحات حسب عامل التمييز. وتحصل المصطلحات التى تظهر كثيرا جدا على عامل تمييز منخفض، ولا تكون مصطلحات تكشف أو بحث جيدة.

ويمكن استخدام المكان location لوزن المصطلحات؛ فالكلمات التى ترد فى العنوان، على سبيل المثال، يحتمل جدا أن تكون قوية الصلة بموضوع الوثيقة، كذلك تلك التى توجد فى مستخلص، إن كان ثمة مستخلص^(١٢). وبدلا من استخدام الكلمات بالشكل الذى ترد عليه، يمكن أن نستخدم جذور الكلمات أو أصولها stemming. وفى عملية البحث، يمكن أن تستخدم التغذية الراجعة للمستفيد لاعطاء وزن أثقل لتلك المصطلحات التى تسترجع الوثائق المتصلة؛ ويمكن أن تضاف مصطلحات من الوثائق المفيدة الموجودة إلى صيغة البحث.

ويمكن الآن أن يعطى توفيق (مزيج) من مصطلحات البحث صيغة بحث لكل استفسار أو سؤال query يمثل التوفيق (المزيج) الموزون لكل من المصطلحات الفردية .

لقد جرب سالتون Salton عددا من هذه الطرق بنجاح في تجارب سمارت SMART، وهي تجسد الآن في الأنظمة العاملة^(٢١). وقد افترض سالتون أن كل مصطلح في وثيقة ما يمكن أن ينظر إليه على أنه متجه Vector؛ وحينئذ فإن إجمالي n من المصطلحات سوف تعطى متجه n-dimensional vector يصف الوثيقة .

ويمكن أن يعامل الاستفسار بالطريقة نفسها، ويضاهي مع المتجهان المتعددا الأبعاد the two multidimensional vectors. وهناك مقياس وجد أنه مفيد وهو معامل ارتباط جيب التمام the cosine correlation coefficient، وهو عبارة عن مقياس للزاوية بين متجهين؛ فإذا تطابقا بشكل كامل، فإن الزاوية بينهما سوف تكون 0°. وإن جيب التمام لـ 0° هو ١. وإذا لم يتوافق المتجهان على الإطلاق، فإن الزاوية بينهما سوف تكون 90°، ويكون جيب التمام 90° صفرا. ومن الناحية العملية، فإن عددا بين صفر و ١ يمكن الحصول عليه لكل وثيقة بالنسبة للاستفسار، وهذا يمكن من ترتيب الوثائق. وتعطى الوظيفة أو الدالة بالمعادلة :

$$\cos (q, d) = \frac{\sum_{i=1}^n q_i d_i}{\left(\sum_{i=1}^n q_i^2 \sum_{i=1}^n d_i^2 \right)^{1/2}}$$

وبدلا من ذلك، يمكن أن نستخدم الشكل الأبسط

$$\text{Strength of association} = \frac{Cab}{\sqrt{(Oa^2 \times Ob^2)}}$$

حيث Oa هي العدد الإجمالي لورود المصطلح a، و Ob هي العدد الإجمالي

لورود المصطلح b، cab هي عدد مرات ورود المصطلحين a و b، الذي اقترحه سبارك جونز Spark Jones^(٢٧).

ومن الأنظمة التي تستخدم طرقاً مشابهة لسمارت نظام NLM CITE^(٢٢)، الذي يقبل الاستفسارات باللغة الطبيعية ويستخدمها كأساس للبحث بعد حذف أي كلمات توقف (يستخدم منها ٦٠٠). ويستطيع النظام أن يعطي مخرجات مرتبة، فهو يستخدم الكلمات الأصلية، ومصطلحات من قائمة رؤوس الموضوعات الطبية: MeSH، ولغة التشفيف المستخدمة في حقل الواصفات، والمصطلحات المترابطة المشتقة من تشغيل الحاسب، وكذلك استخدام توافيق المصطلحات بالصورة التي ألمحنا إليها آنفاً. ويستطيع المستفيد أن يستخدم هذه لكي يصدر أحكاماً أو تقديرات الصلة، والتي يمكن الاستفادة منها لتعديل استراتيجية البحث إذا لم يكن المستفيد مقتنعاً بالنتائج الأولى.

وقد اقترح مارون Maron طريقة مختلفة لحساب الوزن تطبق على كل مصطلح يستخدم لتشفيف وثيقة ما. وباستخدام المصفوفة التي أوردناها آنفاً فإن الوزن الذي يعطى لمصطلح ما هو احتمال أن المستفيد الذي يريد وثيقة معينة Dj سوف يستخدم المصطلح I_i للبحث عنه. هذا الاحتمال w_{ij} يمكن تقديره كما يأتي:

$$\frac{\text{العدد المكتفى ب } Dj \text{ ويستخدم } I_i}{\text{العدد المكتفى ب } Dj}$$

ويرمز إليها على أنها P(I_i|A, Dj) حيث تمثل A فئة المستفيدين كلها. ومع ذلك، فإن هذا الوزن لا يزال يعتمد على تقدير المكشف، في حين أننا يجب أن نستمر في محاولتنا لترتيب الوثائق حسب احتياجات المستفيدين: P(D_j|A, I_i). ويمكننا أن نحول وجهة نظر إلى أخرى من خلال التداول الرياضى:

$$P(D_j|A, I_i) = P(D_j|A) \cdot P(I_i|A, D_j) \cdot c = P(D_j|A) \cdot w_{ij} \cdot c$$

حيث $P(D_j | A_i)$ هي احتمال أن الوثيقة D_j سوف تلبى احتياجات كل المستفيدين بالمكتبة A ، والذي يتم حسابه بنوع الأساليب الاحصائية المنوه عنها آنفاً. وعلى هذا الأساس فإن من الممكن ترتيب الوثائق حسب احتمال أنها سوف تلبى احتياجات المستفيدين^(٢٣).

وتحتاج عملية الترتيب إلى تحسيب أكثر من البحث البولي، ولذلك فهناك حل وسط هو تنفيذ بحث بولي لتقليل عدد الوثائق التي سوف ترتب إلى مائتين مثلاً ثم ترتيب تلك الأخيرة. والطريقة اقترح عملها، وتستخدمها بعض قواعد البيانات، رغم أن الغالبية لازالت تستخدم البحث البولي فقط. وإن مزايا الترتيب ranking كبيرة لدرجة أن المؤكد أننا سوف نشهد تزايداً مستمراً في استعمالها.

الاستدعاء والصلة

إن معظم العمل الذي تم التأسيس مفهومي الاستدعاء والصلة - recall and relevance - قد نفذ على قواعد بيانات صغيرة. وكان الاستثناء هو تقويم مدلاز لعام ١٩٦٥، حيث جرى البحث في قاعدة بيانات تحتوى على ٨٠٠٠٠ راجعاً. وهناك الآن قواعد بيانات كثيرة يحتوى كل منها على ملايين من المراجع. ويصبح من الواضح أن نسبة الصلة التي يمكن أن تكون مقبولة تماماً في مجموعة تتألف من مئات قليلة من المراجع سوف لا تكون مقبولة على الإطلاق في مجموعة تتألف من مليون. وفي تقويم مدلاز، كان متوسط عدد المراجع المسترجعة في البحث الواحد ١٧٥، بنسبة صلة ٥٠٪؛ أى أنه من بين المتوسط العددي ١٧٥ مرجعاً كان هناك ٩٠ غير متصل. وقد أصبح حجم قاعدة البيانات الآن (١٩٩٥) عشرة أمثال عددها في ذلك الوقت؛ وإذا اتبعنا اجراءات البحث نفسها، فسوف نسترجع متوسطاً هو ١٧٥٠ من المراجع لكل بحث، منها ٩٠٠ لن تكون متصلة! لقد كان متوسط نسبة الاستدعاء حوالى ٥٨٪، محسوباً بطريقة تقريبية؛ وكان من الواضح أنه من غير الممكن بحث قاعدة البيانات كلها بالنسبة

لكل بحث لكي نكون أساسا للاستدعاء. فإذا أخذنا البحث المتوسط، وافترضنا أن حوالي ٩٠ مرجعا من تلك التي وجدناها كانت متصلة، مع نسبة استدعاء ٥٨٪. فإذا هذا يعنى ضمنا أننا لا بد قد وجدنا ١٥٥ مرجعا تقريبا، ولكننا أخطأنا ٦٥. ومرة أخرى فإننا إذا طبقنا هذا على قاعدة البيانات الحالية، فسوف يتعين علينا أن نفترض أننا سوف نخطئ عدة مئات من الوثائق المتصلة.

كذلك فإن علينا أن نأخذ في الحسبان عاملا آخر بالإضافة إلى الاستدعاء والصلة : المنفعة Utility، والتي ذكرناها أيضا في الفصل ٢. فإذا نفذنا بحثا باستخدام استراتيجية بحث معينة، فينبغي أن نستدعي بعض الوثائق المتصلة، ونحن نفترض أن هذه الوثائق سوف تكون مفيدة للسائل. فإذا عدلنا البحث لكي نستدعي مزيدا من الوثائق، فمن المؤكد أننا سوف نسترجع بعض الوثائق التي سبق استرجاعها. وإن استرجاعها مرة ثانية ليس مفيدا! وقد تناولت دراسة مدلازر هذه النقطة، حيث كان أحد العوامل التي قيست هو «نسبة الجودة - novel ty ratio»: هل الوثائق المسترجعة جديدة بالنسبة للسائل، أم أنه يألفها من قبل؟ وكما أشرنا في الفصل ٣، فإن استخدام كشاف إسنادي يحتم علينا أن يكون لدينا نقطة بداية تكون وثيقة نعرف من قبل أنها متصلة. ولن يكون نجاحا إذا كشف بحثنا في النهاية الوثيقة التي بدأنا بها - رغم أن هذا إعادة تأكيد على أن استراتيجية بحثنا للبحث صحيحة. وإذا نفذنا بحثا بأي طريقة، فمن المعتاد أن نصل إلى نقطة يصل بنا البحث بعدها إلى إعادة تحديد أماكن الوثائق التي سبق أن وجدناها. وهنا يكون الوقت قد حان لإيقاف البحث، أو أن نتبنى طريقة جديدة بصورة كلية!

ويحتاج جج بلير Blair^(١٦) بأنه مع قواعد البيانات الضخمة التي لدينا الآن، فإن التفكير القديم عن مستويات الاكتفاء بالنسبة للاستدعاء والصلة لم يعد كافيا؛ فالزيادة في الحجم قد أدت لا إلى مشكلة كمية فحسب، وهذا ما أوضحناه في الفقرة السابقة، ولكن أيضا إلى تغيير نوعي في الطريقة التي يجب أن ننظر بها إلى الاسترجاع. إن القدرة على رفض «تفادي dodge» المادة غير المرغوبة تصبح مهمة إلى درجة كبيرة جدا؛ ونحن نحتاج إلى تحقيق نسب صلة أعلى

بكثير، بينما فى ظروف معينة تكون النسب الأعلى للاستدعاء أساسية. والمثال الذى يورده بليز بصفة خاصة هو قاعدة بيانات قانونية، كوثها محاميان لدعم مرافعاتهما فى المحكمة. وقد صاغ المحاميان البحوث بمعاونة اثنين من المساعدين القانونيين، ونفذها اثنان من اختصاصى المعلومات. وكثير من البحوث نفذت فيما وراء ذلك من خلال تداول صياغات البحوث، مع إضافة مصطلحات واستخدام طاقة أو قوة برنامج قاعدة البيانات STAIRS. وقد مررت الوثائق المسترجعة بواسطة البحوث الإضافية إلى المحامين مع تلك التى وجدت من خلال البحوث التى صاغوها. وبصفة عامة، شعر المحاميان أنهما كانا يسترجعان حوالى ٧٥٪ من الوثائق المتصلة بواسطة بحوثهما، واضعين فى الذهن أنهما هما اللذين أسسا قاعدة البيانات بنفسهما. وفى الحقيقة، فإن بحوثهما كانت تجد أن حوالى ٢٠٪ من إجمالى الوثائق المتصلة قد كشفت عنها البحوث المختلفة. أما الوثائق المتصلة الإضافية فقد وجدت فقط عن طريق استراتيجيات بحث موسعة. فى مثل هذا الوضع يكون الاستدعاء العالى والصلة العالية جوهريين حتى يمكن أن نحاجج عن نجاح هذه الحالة. وفى قاعدة بيانات كبيرة، فمن المنتظر أيضا أن نأخذ وقتا أطول لكى نصل إلى النقطة التى لانسترجع فيها وثائق مفيدة. وإن حقيقة أن معظم الشركات القانونية الكبيرة تكون الآن قواعد بيانات لكى تكون سندا لمرافعاتها - تقترح أننا يجب ألا نكون راضين حول نجاح بحث النصوص على الخط المباشر حاليا، حيث أن طرق البحث ينتظر أن تستخدم أكثر بكثير مما كانت عليه فى الأوضاع المعتادة فى الماضى.

تصميم المواجهة

لقد نشر فيكرى وفيكرى مناقشة طويلة ومفيدة جدا عن التصميم العام لمواجهة البحث search interface^(٢٥). وهذه المناقشة تدرس المتطلبات للنظام وأساليب تشغيل الاستفسارات التى يمكن استخدامها لتحقيق تلك المتطلبات. ولناخذ المتطلبات الوظيفية أولا، فنحن إذا واجهتنا حاجة إلى بحث ما، فيجب

أن نؤسس السياق. وسوف يمكننا هذا من اختيار قواعد البيانات والمضيفين hosts (بعض قواعد البيانات متاحة على أكثر من خدمة بيلوجرافية، وقد تكون موجودة أيضا على قرص مدمج). ثم نأخذ تعبير المستفيد عن الاستفسار، والذي قد نحتاج إلى توضيحه عبر المقابلة الشخصية المرجعية المعتادة. ثم بعد ذلك نحتاج إلى دمج أو ترجمة المصطلحات المستخدمة في الاستفسار إلى تلك التي ينتظر أن نجدها في قاعدة البيانات، مثال ذلك : باستخدام لغة مقيدة، لخلق عبارة بحثية. بعدها يمكن أن ننفذ البحث ونحصل على مجموعة من النتائج. وبعد أن نحذف الاصابات المكررة، يمكننا أن نقوم بالنتائج، التي قد يرتبها النظام. هذا الاجراء يطبق على نطاق عام، وهو يصدق على البحث اليدوى كما على بحث الحاسب.

ومن بين أساليب تشغيل الاستفسارات، فسوف نحتاج إلى توحيد disambi-guation معانى مصطلحات البحث، وهذا ممكن باستخدام علاقات المكتز أو سلالمة رتب التصنيف، واستبعاد الكلمات فى قوائم توقف. وقد نحتاج إلى استخدام أصول أو جذور الكلمات لحذف اللواحق. وحيثذ يمكن أن يصاغ الاستفسار فى صورة عبارة بحث بوىلى أو أكثر، والتي قد نحتاج إلى تداولها للحصول على أفضل النتائج. وإن مصطلحات البحث التي لاتعطى النتائج المرغوبة قد تكون أخطاء يمكن مراجعتها مع الهجاء أو الصوت، كما ذكرنا من قبل. وقد يعطينا النظام قيما لصلة اللفظ، بحيث يجعل من الممكن وزن الوثيقة وترتيبها. ويجب أن يكون من الممكن تعديل الاستفسار الأصىلى عن طريق التغذية الراجعة للصلة المبنية على النتائج الأولى.

تصنيف الحاسب

تنزع قواعد بيانات النص الكامل إلى أن تكون ضخمة، الأمر الذى يعنى، كما سبق أن ناقشنا، أنها تسبب مشكلات عند الاستخدام. وقد نكون قادرين على الاستفادة من تشغيل الحاسب لكى يساعد فى تقليل حجم العمل الذى يجب

القيام به لتنفيذ بحث ما . ويمكن استخدام معاملات التشابه التي أشرنا سابقا لاعطاء عناقيد أو جماعات clusters من المصطلحات أو الوثائق للمعاونة في البحث^(٢٦) . ويمكن استخدام عناقيد المصطلحات باعتبارها حواجز أو حدودا hedges، في حين تخدم عناقيد أو جماعات الوثائق لتقليل حجم المجموعة التي يجرى بحثها إجابة لاستفسار معين . ولتتناول طرفي النقيض، في الأول يمكن اعتبار المجموعة كلها (من الوثائق أو المصطلحات) عنقودا أو جماعة cluster واحدة، وفي الثاني يمكن أن ننظر إليها على أنها تتألف من عناقيد بعدد الوثائق أو المصطلحات؛ من الواضح أن كلا الطرفين لايفيد بصفة خاصة في تشغيل البحث، ونحن نحتاج إلى أن نجد قيمة وسطا مناسبة .

وإذا تناولنا المصطلحات، فبوسعنا أن نحسب العلاقات بين أزواج المصطلحات من عدد المرات التي ترد فيها معا في الوثيقة نفسها، على سبيل المثال .

عندئذ يمكننا أن نرتب هذه ونحدد نقطة قطع أو توقف، والتي يمكن اعتبار المصطلحات فوقها مترابطة؛ وتحدد نقطة القطع أو التوقف قوة العلاقة . ويمكننا أن نستبعد المصطلحات التي ترد في وثيقة واحدة فقط، على أساس أن إضافة مثل هذا المصطلح يمكن فقط أن يزيد الاسترجاع بمقدار هذه الوثيقة الواحدة؛ أو تلك التي ترد كثيرا جدا على أساس أن استخدامها لن يكون مفيدا .

(ولنأخذ الطرفين مرة أخرى : المصطلح الذي ورد في كل وثيقة في المجموعة لن يكون له قيمة تمييز!) ويمكن أن نجد أربعة أنواع من الجماعات groups : الخيوط strings، النجوم stars، الزمر cliques، والكتل clumps . وترد الخيوط حينما يكون المصطلح A قوى الارتباط مع المصطلح B مع المصطلح C، وهكذا . وتنزع الخيوط من الناحية العملية إلى أن تكون حلقات loops بسرعة نسبية :

term A → B → C → D → E → A

وتوجد النجوم حينما يكون قوة ارتباط مصطلح ما مع اثنين أو أكثر من المصطلحات متساوية. وترد الزمر حينما تكون مجموعة من المصطلحات مرتبطة كلها معا بالقوة نفسها. والكتل هي شكل ضعيف من الزمر، يكون فيه المصطلح مرتبطا مع واحد أو أكثر من المصطلحات الأخرى فى الكتلة، ولكن ليس بالضرورة مع الجميع. وفى البحث، يمكننا أن نبدأ بمصطلح بعينه ولكننا نجد النتائج غير مرضية؛ وحينئذ يمكن أن نستخدم العلاقات التى سبق تحديدها بين المصطلحات لكى نغير استراتيجيتنا للبحث، كما ذكرنا فى مناقشتنا السابقة عن الحواجز أو الحدود. ومن المعتاد أننا نفكر فى تجميع المصطلحات كما لو كان وسيلة استدعاء، ولكن سبارك جونز أشارت إلى أن هذا يمكن أن يكون وسيلة تحقيق. فإذا بدأنا، مثلا، بأربعة مصطلحات وأجرينا بحثا بوليا فقد لا نسترجع شيئا، ولكن إذا استبدلنا المصطلحات المترابطة فقد نكون قادرين على تحقيق نجاح على مستوى الترابط الذى بدأنا به، وليس بأن نسقط ببساطة مصطلحا أو أكثر للحصول على نتائج على مستوى أقل من الترابط^(٢٧).

وإذا نظرنا إلى العقدة أو التجميع clustering من وجهة نظر الوثائق، فإننا يمكن أن نستخدم النوع نفسه من المعالجة لتحديد أى الوثائق ينتظر أن ترتبط. وبدلا من أن نستخدم أسلوب سالتون لقياس الارتباط بين الوثائق والاستفسارات أو الأمثلة، يمكننا أن نستخدمه لقياس الارتباط بين الوثائق^(٢٨). وعندئذ يمكننا أن نكون عنايد الوثائق، تكون الوثائق فى داخلها مرتبطة ببعضها على مستوى نحدده. بالنسبة لكل عنقود يمكننا أن نحدد متوسطا هو «مركز الثقل أو الجاذبية gravity»، (المركز المتوسط centroid) الذى يمثل العنقود ككل؛ وهذا إما أن يكون وثيقة مخصصة، أو شبه وثيقة محسوبة. فى البحث، يمكننا الآن أن نقصر تشغيلنا على عنقود الوثائق الذى يرتبط مركزه المتوسط ارتباطا وثيقا بالاستفسار. ومن الناحية العملية، يمكننا أن نبني سلالم رتب أو هرميات من العنايد وذلك باستخدام نقاط قطع أو توقف ذات مستويات ارتباط مختلفة؛ وعندئذ يمكننا أن نبدأ البحث عند مستوى يبدو أنه سوف يلبى احتياجاتنا، وهذا

يعتمد على ما إذا كنا نبحث عن تحقيق عال أو استدعاء عال . بالنسبة للتحقيق العالى يمكننا أن نستخدم العناقيد ذات أعلى قيم ارتباط، التى سوف تكون الأصغر بطبيعة الحال؛ وبالنسبة للاستدعاء العالى قد نفضل أن نبدأ بالعناقيد الأكبر وهى التى لها مستويات ارتباط أقل .

ولكى تكون العناقيد المولدة (المنتجة) بواسطة الحاسب مفيدة، فيجب أن تكون ثابتة أو مستقرة إلى حد معقول . فالطريقة التى تعطى عناقيد تتغير تغييرا جوهريا كلما أضفنا إليها وثيقة - لن تكون مفيدة . والاطراد مفيد أيضا؛ فمن المفضل أن يعطى التشغيل عنقودا واحدا، أو يعطى على الأكثر عددا محدودا . وفى حين أن الاطراد ليس ضروريا بصورة مطلقة، من حيث أن العناقيد المختلفة قد تؤدي أداء جيدا بصورة متساوية من الناحية العملية، فإن الثبات أو الاستقرار يبدو جوهريا . ومن المنتظر أن يتحقق هذا حينما تكون قواعد البيانات التى تشغل من أجل العنقدة كبيرة؛ وكما أن خطة التصنيف أو المكتز الذى أعد يدويا سوف يتغير تغييرا جوهريا مع كل وثيقة جديدة تصنف أو تكشف، فى حين أنه سوف يبقى صغيرا، فإنه سوف يصل فى النهاية إلى حالة يصبح فيها متوسط التغيير لكل وثيقة تضاف غير مهم نسبيا . وعلى هذا فسوف يظهر أن قواعد البيانات التى تصلح بصورة جيدة لأساليب العنقدة هى تلك القواعد التى تثبت العنقدة أنها مفيدة معها! والمحاولات الأولى لتطوير أساليب العنقدة نفذت على قواعد بيانات صغيرة وأكبر قاعدة بيانات استخدمت مع تجارب سمارت الأصلية كانت أكثر بقليل من ألف وثيقة . ونحن الآن نملك قدرات تحسب أعظم بكثير متاحة لتنفيذ التشغيل العميق الذى تنطوى عليه أساليب العنقدة، وقد نرى أساليب استبعدت ذات مرة باعتبارها تجريبية خالصة - نراها تصبح عملية واقتصادية فى آن واحد .

حدود مضاهاة الحاسب

من المهم أن نتذكر أن أساليب الحاسب فى البحث أو العنقدة مبنية على

مضاهاة الكلمات حيث أن مجموعات الأرقام خلو من أى محتوى دلالى . وهذا يرى بوضوح أكثر مايرى فى أساليب Soundex لمضاهاة الكلمات عن طريق البتر، والتي تناولناها فى الفصل ١٥، إذ من الواضح أن الخيوط الناتجة ذات الأربعة حروف لامعنى لها فى ذاتها. وهناك عمل يجرى لتطوير نظم إم IR تأخذ فى الحسبان المحتوى الدلالى، ويبدو أن النتاج مرجوة (مبشرة)، من حيث أن من الممكن تحسين الاستدعاء والصلة. ويناقش الفصلان ٦ و ٧ المشكلات التي ينطوى عليها عمل ذلك عقليا، ولكن قد يحدث أن نظم الحاسبات فى المستقبل قد تكون قادرة على محاكاة هذه المعالجة^(٢٩).

النظم الخبيرة (نظم الخبراء)

إن المستفيدين الذين يأتون إلى نظام إم بسؤال هم فى الحقيقة يفتقرون إلى معلومات يحتاجون إليها، ولكنهم ليسوا قادرين على التعبير عن حاجاتهم بوضوح : فلو أنهم عرفوا السؤال، فسوف يكونون على الطريق للوصول إلى الاجابة. وأثناء المقابلة المرجعية يحاول طرف ثالث استيضاح أو استخراج المعلومات من السائل بحيث يوضح السؤال. وفى النهاية الأخرى للعملية، فإن المعلومات التي ترضى السائل قد تكون جاءت من خبير أو اثنين معلى معرفة جيدة بالموضوع.

وقد نكون قادرين على مساعدة السائلين من خلال تطوير نظم حاسبات خبيرة، نخزن فيها المعلومات التي جمعناها من الخبراء مع القواعد والاجراءات التي تمكن المستفيدين من الوصول إلى المعلومات التي يحتاجونها بالرغم من أنهم بدأوا من نقطة الجهل^(٣٠). وعلى هذا فإن النظام الخبير معنى بموازاة الغرض من المقابلة المرجعية، ولكنه يقلل كذلك من خطوات استرجاع المراجع واسترجاع الوثائق عن طريق توفير إجابات مباشرة.

وفى بناء نظام خبير لموضوع بالذات نواجه مشكلات معينة. وأول هذه المشكلات مشكلة جمع المعلومات المتاحة فى نطاق مجال النظام والذي يعرف

بعناية. ويمكن أن يبدأ هذا ببحث الإنتاج الفكرى، ينتظر أن يحدد هؤلاء الذين يعتبرون خبراء فى المجال. والخطوة الثانية هى مراجعة الخبراء أنفسهم، وهنا نواجه المشكلة الثانية. فالخبراء ينبغى أن يكونوا قادرين على تأكيد دقة وكفاية المعلومات التى جمعناها، ولكنهم قد يجدوا أن من الصعب أن يشرحوا كيف طلبوا هم أنفسهم المعلومات. وعلى مر السنين فإننا جميعا نطور عادات عقلية لجمع المعلومات وتشغيلها والتى تمكنا، حينما نواجه مشكلة ما، من الوصول إلى حل بطريقة تجريبية heuristically؛ ونحن نصل إلى قرارات مبنية على الخبرة الماضية دون تعريف لكل خطوة من خطوات عمليات التفكير التى وصلت بنا إلى الإجابة. أما بالنسبة لبرنامج الحاسب فلكى يؤدى وظيفته، يجب أن تكون كل خطوة محددة ومعرفة بوضوح، وإلا فلن يكون البرنامج قادرا على تحقيق العمل الذى أعد من أجله. والمشكلة الثالثة هى أنه عند إيجاد إجابات للأسئلة فإننا لنعتمد فقط على المعلومات المخصصة بالنسبة لهذا الموضوع، ولكننا نستفيد من مدى واسع من المعرفة العامة لتزويدنا بسياق، ومقارنات وقياسات وحالات تساعدنا فى الوصول إلى قرارات تمكنا من الوصول إلى هدفنا. وليس من الممكن عمليا تجسيد كل هذا المدى من المعرفة العامة فى النظام الخبير؛ فلكى نجعل العملية ناجحة فيجب أن نقصر المعلومات التى نضعها فى النظام على تلك التى تكون خاصة بالمجال الموضوعى الذى يغطيه (النظام). ومن ناحية أخرى فإن السائلين قد يضلون خارج هذه الحدود الضيقة، بسبب عدم معرفتهم بها.

وبعد أن نفرغ من تكوين الأساس المعرفى للنظام، ونشرح عمليات اتخاذ القرار التى استخدمها الخبراء فسوف يبقى علينا أن نجسد ما علمناه فى نظام الحاسب، باستخدام واحد من البرامج التى كتبت فعلا لتطوير النظام الخبير، ونصمم مواجهة مناسبة مع المستفيد، واضعين فى الذهن الجماعة المستهدفة التى وضع النظام من أجلها. فإذا ما أنشئ النموذج prototype فيجب أن يختبر، ومن ثم، أن يعدل لتصحيح أى جوانب نقص. وقد أعطى أرديس Ardis

مثالا على الصعوبة التي تظهر في تصميم نظام خبير لمساعدة المستفيدين في بحث عن براءات الاختراع على الخط المباشر؛ وقد كانت واحدة من المشكلات التي لم يتم التعرف عليها أثناء مرحلة التخطيط أن مستفيدين كثيرين لم يعرفوا الفرق بين براءة الاختراع والعلامة التجارية. ومكتيبو المراجع الذين أجابوا عن هذه الأسئلة في العادة لا بد أنهم قد أدركوا ضمنا بطبيعة الحال النوعين من السؤال باعتبارهما مستقلين، وتعين أن يبنى هذا في نموذج بمجرد أنه أصبح واضحا من خلال الإخفاقات مع النظام. ونلخص فيما يأتي خصائص النظام الخبير ومتطلباته:

النظام الخبير:

- ١- يجب أن يمثل المعرفة المخصصة لمجال تخصص الخبير بالطريقة التي يستخدم بها الخبير المعرفة.
 - ٢- يجب أن يجسد عمليات الشرح وطرق تداول عدم التأكد.
 - ٣- يلائم بطريقة نموذجية المشكلات التي يمكن تمثيلها بطريقة رمزية.
 - ٤- يكون أكثر تحملا لأخطاء المستفيد من البرامج التقليدية.
- ولكى يحقق هذا :

- ١- يجب أن يتوفر خبير واحد معترف به على الأقل في المجال الموضوعي.
- ٢- مصادر تجربة الخبير هي الحكم أو التقدير والخبرة.
- ٣- يجب أن يكون الخبير قادرا وراغبا في شرح معرفته / معرفتها.
- ٤- يجب تحديد المشكلة بشكل جيد.
- ٥- يجب أن يحظى مجال المشكلة بإجماع حقيقي.
- ٦- يجب توفير البيانات الخاصة بالاختبار بسهولة^(٣١).

وحتى الآن لا يوجد إلا عدد قليل من النظم العاملة في علم المكتبات والمعلومات. هناك أنظمة لمساعدة مفرس الخرائط، للمراجع، ولتقويم المنح أو التبرعات، ولكن معظمها متخصص جدا. وثمة مجال موضوعي مرتبط بعضه

بالبعض وهو PLEXUS، وقد أعد لعملاء المكتبات العامة الذين يريدون معلومات عن الحدائق^(٣٢). وعلينا أن نتظر لكي نرى أى انتقال جوهرى لأعمال المراجع من البشر إلى الحاسبات، ولكن لاشك أن هذا سوف يكون اتجاهها مستقبلياً، فإن أنصار الذكاء الاصطناعى قد قبلوا الآن الحدود الحالية على عملهم وأخذوا يركزون على ما يمكن تحقيقه.

ملخص

حاول هذا الفصل أن يعطى نظرة شاملة عن استخدام البحث على الخط المباشر، وبعض الخلفيات عن دورها المالى، وأن يدل على نوع الأساليب التى يمكن استخدامها. وإن عدد قواعد البيانات على الخط المباشر ومجالها يعنى أن البحث على الخط المباشر هو الآن الطريقة العادية للحصول على المعلومات بالنسبة لكثير من الناس. وليست كل قواعد البيانات قواعد بيلوجرافية، فهناك قواعد مالية على سبيل المثال، تمكنا من المضاربة على سوق الرصيد، إذا رغبتنا ونحن مرتاحون فى منازلنا! وتشتمل قواعد البيانات بصورة متزايدة على معلومات أخرى غير النص؛ وكما ذكرنا سابقاً، فإن التقانة الآن تسمح لنا باسترجاع الرسوم والصوت (الصوت والصورة). وتشتمل قواعد البيانات الكيميائية على رسوم بيانية بنوية خاصة بالمركبات فى سنوات عدة، ولكن المرسوم هنا بسيطة جداً مقارنة مع ما هو متاح الآن. وتستخدم قواعد البيانات الاحصائية على نطاق واسع؛ وتصبح البيانات الاحصائية الرسمية متاحة للصناعة والتجارة، كما للجمهور العام، وهى أسرع الآن بكثير حيث تجمع بواسطة الحاسب. وإن الطبعة الأولى من Oxford English Dictionary قد احتاجت لجمعها إلى ٤٠ سنة، وهى فى ١٣ مجلداً، وأصبح إصدار ملحق بعد خمس سنوات ضرورياً كنتيجة للتوسع فى عملية التحرير؛ والطبعة الثانية متاحة الآن على قرص مدمج واحد، واحتاجت لإنتاجها إلى ست سنوات. وقد أصبح الأطفال معتادين على استخدام الحاسبات ومصادر الأقراص المدمجة فى المدارس وفى المنازل، وسوف يتوقعون أن يجدوا النوع نفسه من المعلومات

المتاحة لأغراض العمل في حياتهم فيما بعد. وقد رأينا أن اللغات المقيدة التي تجمع بواسطة الجهد العقلي لازالت تستخدم لتحقيق نتائج مرضية. وقد أوضحت تجربة صغيرة أن اللغة الطبيعية قد أعطت تحقياً أعلى واستدعاء أقل من استخدام اللغة المقيدة أو المنضبطة^(٣٣). وإن استخدام اللغتين : الطبيعية والمقيدة قد أعطى المستفيد اختيار إما الاستدعاء العالى أو التحقيق العالى. وسوف ندرس بعض هذه اللغات المقيدة فى الفصول القادمة عن التكشيف المعين.

المراجع

- 1 *Classification research for knowledge representation and organization: proceedings of the 5th International study conference on classification research, Toronto, Canada, June 24–28 1991*, Williamson, N. J. and Hudon, M. (eds.), Elsevier, 1992. (FID 698)
- 2 Austin, C. J., *MEDLARS, 1963–1967*, Bethesda, MD. National Library of Medicine, 1968. The MEDLINE database now contains several million references.
- 3 *Gale directory of databases*, Detroit, Gale Research Inc. 1995. In this edition, v1 lists over 5300 online databases; v2 lists 2015 CD-ROM products and another 2200 databases available on floppy disk, magnetic tape and other media.
Tenopir, C., 'Full-text databases', *Annual review of information science and technology*, **19**, 1984, 215–46.
- 4 Larson, S. E. and Williams, M. E., 'Computer assisted legal research', *Annual review of information science and technology*, **15**, 1980, 251–86.
- 5 Summit, R. K., 'In search of the elusive end user'. *Online review*, **13** (6), 1989, 485–91.
- 6 Cornick, D., 'Being an end user is not for everyone'. *Online*, **13**, March, 1989, 49–54.
Fisher, J. and Bjorner, S., 'Enabling online end-user searching: an expanding role for librarians', *Special libraries*, **85** (4), Fall 1994, 281–91.
Harman, D., 'User-friendly systems instead of user-friendly front-ends', *Journal of the American Society for Information Science*, **43** (2), 1992, 164–74. Suggests that implementing user-friendly front-ends is an inadequate substitute for improving the power of search engines.
- 7 Hartley, R. J., Keen, E. M., Large, J. A. and Tedd, L.A., *Online searching: principles and practice*, London, Bowker Saur, 1990.
- 8 Dalrymple, P. W. and Roderer, N. K., 'Database access systems'. *Annual review of information science and technology*, **29**, 1994, 137–78.
- 9 ANSI Z39.58:1992 *Common command language for online information retrieval*, Bethesda, MD, National Information Standards Organization. 1992.
- 10 Armstrong, C. J. and Large, J. A. (eds.), *Manual of online search strategies*. Boston, Mass., G. K. Hall, 1988.
- 11 'Public access online catalogs'. Markey, K. (ed.). *Library trends*, **33** (4), 1987.

- 523-67. (The point is made here in relation to OPAC searching, but it is of course generally valid.)
- 12 Keen, E. M., 'The use of term position devices in ranked output experiments', *Journal of documentation*, **47** (1), 1991, 1-22.
Keen, E. M., 'Some aspects of proximity searching in text retrieval systems', *Journal of information science*, **18** (2), 1992, 89-98.
 - 13 Walker, S., 'Evaluating and enhancing an experimental online catalogue', *Library trends*, **35** (4), 1987, 631-45.
 - 14 Sievert, M. and Boyce, B. R., 'Hedge trimming and the resurrection of the controlled vocabulary in online searching', *Online review*, **7** (6), 1983, 484-94.
 - 15 Fidel, R., 'Thesaurus requirements for an intermediary expert system', in *Classification research for knowledge representation and organization: proceedings of the 5th International study conference on classification research, Toronto, Canada, June 24-28 1991*, Williamson, N. J. and Hudon, M. (eds.), Elsevier, 1992, (FID 698), 209-13.
 - 16 Blair, D. C., *Language and representation in information retrieval*, New York, NY, Elsevier Science Publishers, 1990.
 - 17 Cleverdon, C. W., 'Optimizing convenient online access to bibliographic databases', *Information services and use*, **4** (1-2), 1984, 37-47.
Cleverdon, C. W. [letter to the editor] *Online review*, **14**, 1990, 35, suggests that intermediaries support Boolean searching because it needs them to make it practical!
Pape, D. L. and Jones, R. L., 'STATUS with IQ: escaping from the Boolean straitjacket', *Program*, **22** (1), 1988, 32-43.
 - 18 Maron, M. E. And Kuhns, J. L., 'On relevance, probabilistic indexing and information retrieval', *Journal of the Association for Computing Machinery*, **7** (3), 1960, 216-44.
Maron, M. E., 'On indexing, retrieval and the meaning of about', *Journal of the American Society for Information Science*, **28** (1), 1977, 38-43.
 - 19 Doszkocs, T. E., 'An associative interactive dictionary (AID) for online bibliographic searching', in *The information age in perspective: proceedings of the ASIS annual meeting, November 1978*, White Plains, NY, Knowledge Industry Publications, 1978, 105-9.
 - 20 Svenonius, E., 'Classification: prospects, problems and possibilities', in *International study conference on classification research, Toronto, Canada, June 24-28 1991*, Williamson, N. J. and Hudon, M. (eds.), Elsevier, 1992, (FID 698), 5-25.
 - 21 Salton, G. (ed.), *The SMART retrieval system: experiments in automatic document processing*, Englewood Cliffs, NJ, Prentice-Hall, 1971.
Salton, G. and McGill, M. J., *Introduction to modern information retrieval*, New York, NY, McGraw-Hill, c1983, Chapter 3.
Salton, G. and Buckley, C., 'Improving retrieval performance by relevance feedback', *Journal of the American Society for Information Science*, **41** (4), 1990, 288-97.
Kantor, P. B., 'Information retrieval techniques', *Annual review of information*

- science and technology* 29, 1994, 53–90.
- 22 Doszkocs, T. E., and Rapp, B. A., 'Searching MEDLINE in English: a prototype user interface with natural language query, ranked output, and relevance feedback' in *Information choices and policies, proceedings of the ASIS annual meeting, 1979*, White Plains, NY, Knowledge Industry Publications, 1980, 131–9.
 - 23 Maron, M. E. and Kuhns, J. L., 'On relevance, probabilistic indexing and information retrieval' *Journal of the ACM*, 7 (3), 1960, 216–44.
Maron, M. E. 'On indexing, retrieval and the meaning of about', *Journal of the American Society for Information Science*, 28 (1), 1977, 38–43.
 - 24 Lancaster, F. W., *Evaluation of the MEDLARS demand search service*, Bethesda, MD, National Library of Medicine, 1968.
 - 25 Vickery, B. C. and Vickery, A., 'Online search interface design', *Journal of documentation*, 49 (2), 1993, 103–87.
 - 26 Van Rijsbergen, C. J., *Information retrieval*, 2nd edn. London, Butterworths, 1979.
 - 27 Sparck Jones, K., *Automatic keyword classification for information retrieval*, London, Butterworths, 1971.
Needham, R. M. and Sparck Jones, K., 'Keywords and clumps: recent work on information retrieval at the Cambridge Language Research Unit', *Journal of documentation*, 20 (1), 1964, 5–15. Included in *Theory of subject analysis . . .*
 - 28 Salton, G. and McGill, M. J., *Introduction to modern information retrieval*, New York, NY, McGraw-Hill, c1983, Chapter 6, section 4.
 - 29 Sembok, M. T. and van Rijsbergen, C. J., 'SILOL: a simple logical-linguistic document retrieval system', *Information processing & management*, 26 (1), 1990, 111–34.
 - 30 Poulter, A., Morris, A. and Dow, J., 'LIS professionals as knowledge engineers', *Annual review of information science and technology*, 29, 1994, 305–50.
Vickery, B. C., 'Knowledge representation: a brief review', *Journal of documentation*, 42 (3), September 1986, 145–59.
Alberico, R. and Micco, M., *Expert systems for reference and information retrieval*, Westport, CT, Meckler, 1990.
Artificial intelligence and expert systems: will they change the library?, Lancaster, F. W. and Smith, L. C. (eds.), Urbana-Champaign, University of Illinois Graduate School of Library and Information Management, 1992. (Clinic on library applications of data processing: 1990)
 - 31 Ardis, S. B., 'Online patent searching: guided by an expert system', *Online*, 14 (2), March 1990, 56–62.
 - 32 Vickery, A. *et al.*, 'A reference and referral system using expert system technique', *Journal of documentation*, 43 (1), March 1987, 1–23.
 - 33 Rowley, J. E., 'A comparison between free language and controlled language indexing and searching', *Information services and use*, 10 (3), 1990, 147–55.

التكشيف المعين ١ :

الدلالات

تناولنا في الفصل ٣ الطرق التي يمكن بواسطتها أن تستق الكشافات المطبوعة من المعلومات التي تعرفها الوثيقة. وفي الفصل ٥ تناولنا بعض الطرق التي يمكن بواسطتها بحث الملفات على الخط المباشر، ومرة أخرى باستخدام المعلومات المعروضة في الوثيقة، مثال ذلك: في العناوين، أو في المستخلصات أو في النص الكامل؛ وقد أشارت المناقشة إلى بعض المشكلات التي ينتظر أن تظهر عند عمل ذلك، وقد أشرنا عند مرورنا إلى استخدام «اللغات المقيدة»، للمساعدة في حل تلك المشكلات، دون أن نبين في ذلك الوقت ما الذي تعنيه اللغة المقيدة. وقد رأينا أيضا في الفصل ٢ أن بحث النص الكامل يعطى أعلى مستوى ممكن من الشمول، الذي ينزع إلى أن يرتبط باستدعاء عال ولكن مع صلة منخفضة؛ وقد نرغب في أن يكون لدينا طريقة ما للتلخيص لكي نكمل التكشيف العميق لبحث النصوص. وإن مناقشة هذه المشكلات تقود إلى التكشيف المعين : assigned indexng .

يجب أولا أن نختار الكلمات التي سوف نستخدمها في بحث النظام وذلك بمحاولة التفكير في كل الكلمات التي قد يكون مؤلفو كل الوثائق التي كشفناها استخدموها لوصف الموضوع الذي نهتم به، وإذا ما اخترنا الكلمات، فيجب أن نفكر في الأشكال المختلفة التي قد تكون وردت عليها. ويستخدم البتر كوسيلة لدمج الأشكال المختلفة للكلمات، ولكن ليس دائما؛ فمثلا سوف تسترجع *TEACH كلا من : teacher و teaching ولكن ليس taught. وثانيا، سوف تحتاج غالبا إلى البحث عن توافيق الكلمات؛ أزواج الكلمات أهم من

الكلمات الفردية وحدها، ولكننا نجد أنفسنا غالبا راغبين فى ربط أكثر من كلمتين. عملية الربط هذه، كما رأينا، هى عملية تقاطع بين الأقسام-Class inter-section. ولنستخدم مثلا سبق أن أوردناه، وهو أن مجموعتنا من الوثائق (عالم الموضوعات) تحتوى على فئة من الوثائق تحتوى على كلمة «ماء water»، وكل فئة من هذه الفئات تكون قسما، وإذا كنا نبحث عن وثائق عن «تلوث الماء water pollution» فإننا نبحث عن تقاطع هذين القسمين. وأية طريقة تتضمن تقاطع الأقسام سوف تكون طريقة قوية فى ضغط العدد الإجمالى للوثائق المسترجعة؛ وقد رأينا كذلك أن ضغط الاستدعاء يصاحبه غالبا تحسين فى الصلة؛ وعلى هذا فإننا سوف نتوقع أن يكون الربط Coordination طريقة مفيدة للحصول على صلة محسنة. ومن ناحية أخرى، فإن توحيد القسم ('A' OR 'B') يزيد العدد الإجمالى للوثائق المسترجعة، وعلى هذا يمكننا أن نتوقع أن يكون تضمين مصطلحات بديلة وسيلة لتحسين الاستدعاء.

وقد أوضحنا أيضا أن Water وحدها قد لا تسترجع كل الوثائق التى نهتم بها، لأنه قد تستخدم مصطلحات مختلفة ولكنها مرتبطة: البحر Sea، والنهر river، وقد أشرنا فى الفصل ٥ أن مضاهاة الحاسب لاتتضمن المحتوى الدلالى، وعلى هذا فإنها لا يمكن أن تؤدى إلى المصطلحات المترابطة مباشرة؛ وحتى تصنيف الحاسب مبنى على عوامل مثل الورد المتلازم، وليس على تشابه المعنى. ولكى ننفذ بحثا كافيا لمجموعتنا من الوثائق، فيجب علينا أن نفكر لا فى الكلمات التى كنا مهتمين بها فقط، ولا فى جميع الأشكال التى سوف تستخدم عليها فحسب، ولكن أيضا فى كل الأشكال البديلة أو المترابطة. وحينئذ يكون علينا أن نقرر كيف نربط هذه الكلمات لكى نسترجع الوثائق المتصلة، فى حين نستبعد فى الوقت نفسه الكلمات أو توافيق الكلمات التى سوف تسترجع مادة غير متصلة. ومن الواضح أن هذه عملية معقدة تماما، وإذا أردنا أن نؤديها بشكل جيد فإننا نحتاج إلى شىء من الارشاد: قائمة بالكلمات التى تبين علاقاتها وتدل على الطرق التى من خلالها يمكن توفيقها معا بطريقة مفيدة لاعطاء التقاطع بين

الأقسام التي نهتم بها. ومع ذلك، فقد أشرنا في الفصل ١ إلى أن ما نحاول عمله فعلا هو أن ننفذ عملية مضاهاة بين الرسائل الى هي في شكلها غير المشفر (المكود) تعد المخرجات لنظامنا والرسائل - في شكلها غير المشفر أيضا - التي تمثل الأسئلة التي نقدمها للنظام. وإن مفهوم المضاهاة هذا يدعمه بطبيعة الحال وبقوة فحوصنا للنظم المبنية على الحاسب، التي تعتمد على الحاسب في مضاهاة الكلمات التي تحتوي عليها أسئلتنا مع الكلمات التي توجد في الوثائق.

والآن إذا كنا سوف نستخدم قائمة بالكلمات لمساعدتنا في بحثنا، فسوف يظهر أننا سوف نزيد من فرص تحقيق عمليات مضاهاة ناجحة إذا استخدمنا قائمة الكلمات نفسها لفك شفرة الوثيقة في مرحلة المدخلات، وعيناً assigned الكلمات المناسبة للوثائق بأنفسنا دون أن نعتد على اختيارات المؤلفين. بمعنى آخر نحن نكون لغة تكشيف ونستخدمها لعمليتي فك الشفرة معا: المدخلات والسؤال. ويطلق على هذه الأنظمة: أنظمة التكشيف المعين، وقد خصص معظم ما تبقى من هذا الكتاب لمشكلات إنشاء واستخدام مثل هذه الأنظمة. وفي هذا الفصل سوف ندرس بعض المشكلات النظرية الأساسية^(١).

اختيار المصطلحات

يعرف التكشيف المعين باسم تكشيف المفاهيم Concept indexing، لأن مانحاول عمله هو إفراد المفاهيم الموجودة في كل وثيقة. (المفهوم Concept: فكرة عن قسم من الأشياء؛ فكرة عامة)^(٢).

ويقترح أحد التحليلات أن هناك خمس فئات من المفاهيم: الوجودات enti- ties؛ النشاطات activities؛ المجردات abstracts؛ الخواص properties؛ المتغايرات heterogeneous. ويعبر عن المفهوم مصطلح term قد يتألف من أكثر من كلمة. (المصطلح: كلمة أو تعبير له معنى محدد في بعض الاستعمالات، أو قد يكون خاصا بعلم ما، أو فن، أو مهنة أو موضوع)^(٣). ويمكن أن نتناول كل واحدة من هذه الفئات بتفصيل أكثر. الوجودات هي الأشياء التي تعطى

معنى دلالية denotative،^(٤) أى أننا يمكن أن نفردها بأن نشير إليها. وقد تكون مادية فيزيائية physical، مثل المادة matter، أو الظواهر الفيزيائية؛ أو قد تكون كيميائية، مثل حالات الجزيئات molecular، المعادن؛ وقد تكون حيوية، مثل الكائن الحي؛ أو قد تكون نتاجا مصنعا (artefacts) أى المواد المصنعة. والنشاطات يعبر عنها عادة بالأسماء الفعلية، مثل: lubricating و building، رغم أننا فى بعض الحالات نجد صيغة المبنى للمجهول وليس المبنى للمعلوم، مثل lubrication. وتحليل المجردات عادة إلى النوعيات qualities أو الحالات، وتعطى معانى مفهومية connotative، أى أن كل واحد منا قد يعطى لها معنى مختلفا وهذا يعتمد على حصيلتنا من الخبرة. وقد تكون فيزيائية؛ مثل الطاقة؛ أو رمزية مثل العدالة كشخص معسوب العينين؛ أو سلوكية، مثل الصدق (الذى كان تعريفه مشكوكا فى مناسبة واحدة ملحوظة على الأقل). والخواص من نوعين، يميزهما الشكل النحوى grammatical. الأشكال النعتية يمكن أن تستخدم مع الاسم فقط، والذى تصفه بطريقة نعتية attributive أو ذاتية subjective، ومرة أخرى تعطى معنى مفهوما. قد تتصل بالبصر مثل معتم، مشرق، متناسق؛ أو بالصوت مثل مرتفع، موسيقى؛ أو بالحواس الثلاث الأخرى: اللمس، والذوق والشم. وقد تتصل أيضا بالخواص الآلية (الميكانيكية)، مثل مرن، جامد. والأشكال الإسمية تصف الخواص المادية التى يمكن قياسها، مثل الجمود، الانعكاسية، الارتفاع (فى الصوت). وسوف يكون واضحا على الفور أنه سوف تكون هناك فى حالات كثيرة علاقة محددة بين النوعين. فيمكننا أن نشير إلى جمود قضيب من الحديد، مثلا، وفى هذه الحالة فنحن نفكر فى خاصية property؛ وقد نشير إلى قضيب جامد، وفى هذه الحالة فإننا نستخدم الخاصة لتعريف نوع الموجود الذى نتناوله.

وتكوّن المفاهيم المتغايرة حقيقية شديدة الاختلاط، من حيث أنها تمثل عادة مفاهيم يمكن تحليلها إلى اثنين أو أكثر من المفاهيم الأبسط التى يمكن أن تدرج

في الفئات الأخرى، ولكنها مع ذلك لا تعتبر مفاهيم موحدة وتعامل كما هي .
وقد اقترح ويليتس Willets⁽⁵⁾ بعض الأنواع .

أدوار الانسان Roles of man (الموجود + النشاط، الموجود + الخاصة) مثل
المدرس، مالك الأرض .

جماعات الانسان Groups of man (الموجود + المجرد) مثل المجتمع،
مؤتمر أنواع المباني (الموجود + النشاط + الخاصة) مثل مكتبة، مسرح العلم أو
الموضوع Discipline (الأربعة جميعا) مثل الفيزياء، الطب، جماعات الكيماويات
(الموجود + النشاط، الموجود + الخاصة) مثل الحفازون،
البوليمرات polyomers .

وينظر أوستن Austin⁽⁶⁾ إلى الجماعات على أنها تجميعات aggregates، في
حين أن الباقية يمكن أن تدخل في معظم نظم الكشف بسهولة تامة كما هي .
هل هناك مزايا تتحقق بمزيد من التحليل؟

في خمسينات ق ٢٠ عمل فريق في جامعة Case Western Reserve على نظام
يعرف بتحليل العوامل الدلالية Semantic Factoring⁽⁷⁾ . كان الهدف هو تجزئ
كل مفهوم إلى مجموعة من المفاهيم الأساسية يطلق عليها العوامل الدلالية .
وبسبب طبيعتها الأساسية، فسوف لا يوجد من هذه العوامل إلا عدد
محدود . والمصطلح قد يُدَلُّ عليه بواسطة توفيق مناسب من العوامل الدلالية،
وقد مكنَّ استخدام مجموعة مركبة من الأدوار roles والروابط links مكن
المكشف من أن يكتب «مستخلصا تلغرافيا» يمثل موضوع الوثيقة في ملف
الحاسب .

ومن الواضح أن الطريقة قوية، ولكنها عرضة لبعض الشكوك والاعتراضات .
إلى أى مدى بالضبط يمكن للمرء أن ينفذ مثل هذا التحليل؟ على سبيل
المثال : Heat and temperature يمكن تخصيصهما على النحو الآتي :
movement of molecules . ومرة أخرى، من الممكن تخصيص مفهوم عن طريق استخدام

بعض صفاته؛ أوروبما كان الأهم من ذلك، هل يمكن دائما تخصيص كل الصفات الخاصة بمفهوم معين؟ مثال ذلك الترمومتر يمكن تخصيصه على أنه instrument : measuring : temperature آلة : تقيس : الحرارة، والبارومتر على أنه instrument : measuring : pressure آلة : تقيس : الضغط . ولايكشف أى منهما حقيقة أنهما قد يشتركان فى عوامل أخرى، مثال ذلك : حقيقة أنهما وسائل تحتوى على الزئبق - فى - زجاج . ومن المؤكد أنه بالنسبة لمعظم الأغراض يشترك البارومتر الزئبقى مع البارومتر اللاسائلى aneroid بأكثر مما يشترك مع الترمومتر، ولكن قد لا تكون هذه هى الحالة إذا كنا نفكر فى صانع الآلة . فإذا بدأنا نفكر فى فرد معين، فقد لا نجد صعوبة فى أن نضع اسما على الهدف من أفكارنا؛ قد نجد أنه من المستحيل فى الظروف العملية أن نفكر فى كل المصطلحات الممكنة التى يمكن أن نحتاجها لتخصيص فرد ما بدون تسميتها . الجنس، العمر، الجنسية، الوضع العائلى، الحالة الاجتماعية، الطول، الوزن، الوظيفة، اللغة، الدين - قائمة لانهاية لها . وعلاوة على ذلك، قد نجد أنفسنا فى وضع لانعرف فيه كل المعلومات التى نحتاجها؛ يجب أن نتذكر أننا نتعامل مع معلومات فى مجموعة من الوثائق، وأن هذه سوف تكون غير كاملة فى العادة .

لدينا كذلك مشكلة فى تحليل مفاهيم معينة تفقد دلالتها إذا فككت إلى أجزائها المكونة . فمثلا soap opera ليست نوعا من الأوبرا، كما أنها ليست نوعا من الصابون^(٨)؛ و«لحظة صدق moment of truth» لايمكن تحليلها أبعد من ذلك؛ والطائر الأسود هو نوع مخصص من الطائر، ولكن هناك طيوراً سوداء كثيرة^(٩)؛ والبطة المطاطية a rubber duck ليست نوعا من البط (رغم الشاهد الذى يتمثل فى بعض إعلانات التليفزيون) وفى أيامنا هذه نادرا ما تصنع من المطاط^(١٠) . ولحسن الحظ، فإنه رغم أن المشكلات النظرية المتضمنة لم تحل جميعا، فإنه يمكن، من الناحية العملية إيجاد الحلول الناجعة بصورة معقولة .

اختيار شكل الكلمة

ينبغي أن يكون قد أصبح واضحاً أثناء مناقشتنا السابقة لفئات المفاهيم أن - مع استثناء وحيد هو الخواص النوعية التي لا يمكن أن تقف وحدها - كل المفاهيم المتضمنة قد دُلَّ عليها بواسطة الأسماء. وحتى النشاطات أشير إليها بواسطة أسماء الأفعال سواء المبنية للمعلوم أو للمجهول، مثل الفهرسة والتصنيف. وفي الحقيقة، فإن المعيار في لغات التكشيف هو استخدام الأسماء بقدر المستطاع، وقد وضعت مجموعات مختلفة من القواعد لكي ترشد إلى استخدام المفرد والجمع. وقد انبنى جدول ٦,١ على القواعد التي أعطاها:

EJC Thesaurus of engineering and Scientific terms

الذي وصفناه في الفصل ٢٥، في حين أن ISO^(١١)، و BSI^(١٢)، و ANSI^(١٣) قد نشرت كلها معايير قياسية لإنشاء المكانز. ومن القواعد المفيدة للإشارة of thumb:

how much ? - use the singular; how many ? - use the plural

إلى أى مدى ؟ - استخدم المفرد؛ كم عدد؟ - استخدم الجمع.

جدول ٦,١ اختيار شكل المفرد أو الجمع من الاسم

| Type of term | Use singular | Use plural |
|---|---|--|
| Materials Properties | When specific, e. g polythene density | When generic, e.g plastics chemical properties |
| Objects Events Objects specified by purpose | | Cars Laws Wars Lubricants |
| Processes Proper names Disciplines Subject areas | Lubricating Earth (the planet) Law War | |

المتجانسات

يستخدم الهمجاء نفسه أحيانا لكلمات مختلفة، فى حين قد تنطق وقد لاتنطق النطق نفسه، مثل Sow و Sow، China و China. وقد ينشأ هذا عن التشبيه أو الاستعارة مثل الكناية، أو المجاز المرسل Synecdoche، والذي نستخدم فيه جزءا من الوصف لنعنى الكل؛ وقد يكون من خلال المشابهة أو القياس Analogy، حينما تستخدم مصطلحات "Filter" مثل من مجال الهندسة الهيدروليكية بواسطة المهندسين الكهربائيين؛ أو قد تكون ببساطة عرضا إيتيمولوجيا etymological accident. وأيا ما كان السبب، فإن من المنتظر أن يحدث ليس إذا لم نفعّل شيئا لتمييز هذه الكلمات. وإن إحدى الطرق لعمل هذا هى أن نصف كل واحدة بكلمة أخرى بين قوسين لتوضيح السياق ومن ثم المعنى، مثل:

PITCH (Bitumen)

PITCH (Football)

PITCH (Music)

PITCH (Slope)

وإذا لم نميز المتجانسات homographs فسوف نحصل على صلة منخفضة؛ وسوف تعتمد خطورة هذا على تغطية نظامنا. فعلى سبيل المثال، لو أن نظامنا يغطى الموسيقى فقط فلن يكون هناك مشكلة مع الكلمة pitch، حيث أن المعانى الأخرى غير الموسيقية لاينتظر أن تظهر فى مرحلة المدخلات، ومع ذلك فقد أشرنا^(١٤) إلى أن الأسماء الانجليزية العشرين التى ترد أكثر من غيرها كل منها له سبعة معان، ولهذا فمن الواضح أننا يجب أن نكون على دراية بالمشكلة.

العلاقات

رأينا أنه بالإضافة إلى اختيار المصطلحات والشكل الذى يجب أن تستخدم

عليه، فإن هناك نوعين من العلاقة بين المصطلحات يجب أن نضعهما في الحسبان: تعرف المصطلحات التي تدل على الموضوعات المترابطة، مثل الماء، والبحر والنهر، وربط المصطلحات غير المترابطة أصلا لتكوين الموضوعات المركبة. وهناك موضع يمكننا فيه أفراد أنواع المصطلحات التي يستخدمها المؤلفون، وكيف يتم ربطها وهو العناوين التي يعطونها لأعمالهم. وإذا درسنا جدول ٦,٢ بعناية، فسوف نرى أولا وقبل كل شيء أن العناوين، التي أخذت من فهرس مكتبة الكونجرس، تقع في ثلاث جماعات رئيسية: التربية، والزراعة، وكتب الطبخ. وفي داخل كل مجال من هذه المجالات الموضوعية يمكننا أن نرى أمثلة على النوعين من العلاقة، ويمكن أن نرى أن نوعا منهما دائم، وينشأ من تعريفات الموضوعات المتضمنة، في حين أن النوع الثاني ينشأ من الارتباطات التي نجدتها في الوثائق، ويمثل ارتباطات مؤقتة، لغرض خاص. ويعرف الأول بالعلاقات الدلالية Semantic relationships؛ القمح Corn هو دائما نوع من الحبوب cereal. ويطلق على الثاني العلاقات التركيبية Syntactic: القضاء على الآفات Disinfestation هو نشاط ينفذ على المحصول، في هذه الحالة القمح grain، وهو موجود؛ والمهاجرون الملونون هم ناس، موجودات، يتم تعليمهم، نشاط، في بريطانيا، مكان. وهذا يقترح أن لغتنا للتكشيف يجب أن تحتوى على بديل للمعجم، لكي يبين العلاقات الدلالية، وعلى قواعد نحوية، لكي تعالج العلاقات التركيبية. في بحث الحاسب، يمكن أن يكون النحو أو قواعد اللغة توفيرا أوليا للمنطق البولي، ولكن في الكشافات المطبوعة أو ترتيب الرفوف فإننا قد نرغب في بيان العلاقات الأكثر تعقيدا.

العلاقات الدلالية

نجد أن هذه العلاقات يمكن تناولها في ثلاث جماعات: التقابل equivalence، والرتبية hierarchical، والقراءة/ النسبة affinitive/ associative. والجماعتان الأوليان مباشرتان بدرجة معقولة، ولكن تعريف الجماعة الثالثة أقل وضوحا بكثير، وهي التي تسبب معظم المشكلات عند التطبيق.

ويمكن أن ندرس الأنواع المختلفة لكل فئة من الفئات الثلاث بشيء من التفصيل، وقد أخذنا معظم أمثلتنا من جدول ٦,٢ .

جدول ٦,٢ الموضوعات المترابطة

Concept analysis

- 1 Education of women in India 1921–1966.
 - 2 Acceleration and the gifted.
 - 3 The costs of education.
 - 4 The teaching of Physics at university level.
 - 5 Teaching French: an introduction to applied linguistics.
 - 6 Saga of the steam plow. [plough]
 - 7 The main course cookbook.
 - 8 The corn earworm in sweet corn: how to control it.
 - 9 Wheat.
 - 10 The potato.
 - 11 New first year mathematics: teacher's book.
 - 12 Radiation disinfestation of grain.
 - 13 The education of coloured immigrants in Britain.
 - 14 Modern corn production.
 - 15 The elementary school: a perspective.
 - 16 Agricultural financing in India.
 - 17 Technology of cereals.
 - 18 Meat, fish, poultry and cheese . . .
 - 19 A cyclopedia of education. [i.e. encyclopedia]
 - 20 Soups and hors d'oeuvres.
 - 21 Curriculum theory.
 - 22 A world of nut recipes from soups to savories.
 - 23 Economic aspects of higher education.
 - 24 The pecan cookbook.
 - 25 Education improvement for the disadvantaged in an elementary setting.
 - 26 The evolution of the comprehensive school.
 - 27 The world book of pork dishes.
 - 28 New media and college teaching.
 - 29 Potatoes in popular ways.
 - 30 Educational aids in the infant school
 - 31 The planetarium: an elementary School teaching resource.
 - 32 Vegetable cookbook.
 - 33 English in the primary school.
 - 34 Talking about puddings.
-

التقابل

المترادفات والأضداد

أشباه المترادفات

المتصل نفسه

التداخل

الهجاء المفضل

الاختصارات، الكلمات الاستهلاكية

المصطلحات الجارية والمستقرة

الترجمات

إن اللغة الانجليزية هي لغة غنية بالمترادفات وأشباه المترادفات، لأن جذورها تمتد إلى اللغة التوتونية واللغة الرومانسية. (الناشئة عن اللاتينية) وفي حين أن من الصحيح أن قصيدة Wordswovrh الغنائية قد لا تكون مؤثرة مثل كل من:

Intimations of immortality و Hints of Deathlessness، فإن الأولى هي صياغة صحيحة للموضوع مثل الأخيرة وكثير من الموضوعات لها اسم عام واسم علمي: البطاطس potato و Solanum tobersum؛ والاستعمال الأمريكي يختلف عن البريطاني أو الاسترالي: primary school و elementary school؛ والمؤلفون يختلفون في استعمالهم: الكلمة College تستخدم بأكثر من معنى. فإذا لم ندمج المترادفات، فإننا بذلك نفصل أدب القارئ العام عن أدب الخبير، ونفصل الأمريكي عن البريطاني، ونفصل مؤلفا عن آخر؛ وهذا من شأنه أن يزيد من الصلة على حساب الاستدعاء. فإذا أدمجنا المترادفات، فمن المتظر أن نزيد الاستدعاء على حساب الصلة.

وقد يبدو غريبا أن نضمن الأضداد antonyms مع المترادفات Synonyms، ومع ذلك ففي محاولتنا لاسترجاع المعلومات قد نجد من المفيد غالبا أن نعاملها بالطريقة نفسها. أما من الناحية العملية، فإن الأطفال الموهوبين هم غالبا محرومون! وتمثل أشباه المترادفات غالبا نقاطا على المتصل نفسه، أو تداخل المفاهيم. وقد تتداخل الأضداد وأشباه المترادفات؛ الخشونة والنعومة: roughness و Smoothness قد يظن أنهما متضادان، ولكنهما يقعان في المتصل continuum الذى يمثل غالبا حكما ذاتيا. وقد نجد صعوبة في أن نميز بوضوح مرحلة ما قبل المدرسة، ومدرسة الأطفال infant school والمدرسة الابتدائية primary؛ والمدارس الشاملة comprehensive تتداخل مع المدارس الثانوية Secondary؛ والاقتصاد economics والتكاليف والتمويل Financing لا يمكن التمييز بينها بوضوح، والمسافة بين نقطتين هي الطول. والحالات الأربع الأخرى واضحة بذاتها:

plow (US) = plough (UK)

labour (US) = labour (UK)

ERIC = Educationd Resources Information Clearinghouse

Third World, Developing Countries, Underdeveloped Countries, Zhurnal = Journal.

وتتضمن علاقة التقابل أنه سيكون هناك أكثر من مصطلح واحد للدلالة على المفهوم نفسه. وفي اللغة المقيدة أو المحكومة (المنضبطة) فإن المعتاد أن نختار مصطلحا واحدا ونعتبره المصطلح المفضل preferred term، ونستخدم هذا المصطلح الواحد في تكشيفنا. ويجب بطبيعة الحال أن نحتاط فنوفر للمستفيدين الذين يبحثون عن المعلومات تحت واحد من المصطلحات الأخرى، وقد ناقشنا هذا بعد في الجزء الذى يوضح العلاقات الدلالية.

الرتبية

الجنس - النوع

الكل - الجزء

إن النوع المعتاد من العلاقة الرتبية Hierarchical هو علاقة الجنس Genus بالأنواع species، التي تتضمن علاقة اشتمال القسم (كل أ هي ب؛ بعض ب هي أ). وهي أكثر ما تكون وضوحا في العلوم البيولوجية (علوم الحياة) (كل الثدييات mammals فقاريات vertebrates؛ بعض الفقاريات ثدييات)، ولكنها موجودة أيضا في مجالات موضوعية أخرى؛ وفي الحقيقة فإن قدرا كبيرا من التصنيف معنى بتكوين سلالم الرتب hierarchies. ويميز أوستن Austin ما يطلق عليه العلاقات شبه الشاملة عن العلاقات الشاملة الحقة: quasi - generic from true generic، باستخدام قانون الدوام permanence؛ البطاطس هي دائما نبات من النوع Solanum tuberosum، ولكنها تظهر في طبق الغداء الذي نتناوله كجزء من الوجبة، أو قد يستخدمها الأطفال لطبع تصميمات بسيطة. والقبة السماوية planetarium هي نوع من مصادر التعلم. لحم البقر Beef ولحم العجل veal هما نوعان من اللحم.

أما علاقات الكل - الجزء فهي ليست شاملة. فالعجلة wheel ليست نوعا من الدراجة، كما أن الباب ليس نوعا من المنزل. ومع ذلك فإن من المناسب أن ننظر إلى علاقات الكل - الجزء باعتبارها علاقات رتبية، وقد اقترح أن يميز النوعان باعتبارهما: أصلى أو وراثي Genetic ومجزئ أو تبعيضي Partitive^(١٢). ونوضح العلاقة التبعية من خلال أمثلة أربعة خاصة:

أ أجهزة الجسم وأعضاؤه

ب الأماكن الجغرافية

ج العلوم أو مجالات الموضوعات

د البنى الاجتماعية الرتبية

في كل حالة، ينبغي أن يدل اسم الجزء على اسم الكل بصرف النظر عن السياق، بحيث يمكن أن تنظم المصطلحات على أنها سلالم رتب منطقية.

علاقات القرابة / النسب

الربط (التساوي في الرتب)

الوراثية (الأصلية)

المتعاصر أو المتزامن

السبب أو الأثر

الآلات

المواد

التشابه

لأن هذه العلاقات هي الأقل وضوحا في التعريف، كما أنها ليست واضحة وضوحا مباشرا، فهي الجماعة التي تسبب معظم المشكلات في لغة التكشيف. وفي الحقيقة فإن كوتس Coates^(١٦). قد انتقد قرمك LCSH لأنها تتضمن هذه العلاقات بشكل يبدو أنه طريقة عشوائية تماما. (وقد اقتصر العلاقات الجديدة في داخل قرمك، حديثا، على علاقات التقابل والعلاقات الرتبة). ورغم الصعوبات، فينبغي أن نقوم بمحاولة التعامل مع هذه العلاقات من خلال إدراك أنها أولا وقبل كل شيء موجودة، ثم نحاول تحديد ذاتيتها بطريقة مقننة.

بعض هذه العلاقات يقدم مشكلات أقل من غيره. فالربط أو التساوى في الرتبة هو في الحقيقة منتج جانبي (عرضي) لعلاقة الجنس: الأنواع المندرجة تحت الجنس نفسه متساوية. وعلى هذا فإن القمح والحنطة Wheat, corn هما معا فرعان من محصول الحبوب؛ ويمكن النظر إلى كل الأصناف التالية على أنها حلقات متتابعة للوجبة: و-savi [u] ries, hors d' oeuvres, soups, entrees, main-courses, puddings. ومما يجدر ملاحظته أنه إذا كان التقسيم من هذا النوع ثنائيا dichotomous، أى إلى أ وأ، فإن النتيجة ستكون إعطاء مفهومين متضادين، مثل الذكر والأنثى، الشعر والنثر. ولهذا السبب فإن الأضداد تعتبر أحيانا واقعة مع جماعة القرابة وليس مع جماعة التقابل.

والعلاقات الوراثية Genetic أو الأصلية هي أيضا مباشرة، مثل الأم - الابن؛ فهنا أيضا يمكن أن نلاحظ أن المستوى الأول من التقسيم الوراثي أو الأصلي سوف يعطى مفاهيم متساوية في الرتبة، مثل الابن - الابنة. والمتعاصر

(المتزامن) concurrent يحيل إلى نشاطين يحدثان في الوقت نفسه ويكونان مرتبطين، وهكذا فهما محتاجان لتفسير أوسع؛ مثال ذلك: التربية - التدريس. والسبب cause والأثر effect هما أسهل في تحقيق ذاتيتهما، رغم أنهما كانا بطبيعة الحال موضوعا لحديث فلسفي مطول؛ ومن الأمثلة المتفائلة من مسودة المقياس البريطاني الموحد British Standard: التدريس - التعليم - teaching Learning وقد حل محل هذا في النسخة النهائية مثال آخر هو: diseases - patho- gens. والآلات instruments، مثل teaching - media، والآلات، - plasticfilm transparencies، هي في العادة واضحة إلى حد كبير.

والفئة الأخيرة التشابه similarity، ربما كانت الأصعب في علاقات القرابة، من حيث أنها تتضمن بالضرورة حكما ذاتيا؛ إلى أي حد يكون مفهومان متشابهين بالنسبة لنا بحيث يمكن التعرف على العلاقة؟ ينبغي ألا نتوقع أية درجة كبيرة من الاطراد بين لغات التكشيف المختلفة.

ومن المهم أن نقارن أنواع العلاقات التي ناقشناها هنا مع تقسيم فنوى مشابه للعلاقات كشفت عنه الارتباطات النفسية واللفظية. وقد يحدث أن دراسة مثل هذه الارتباطات قد تلقي مزيدا من الضوء على أنواع العلاقة التي تحتاج للتعامل معها في لغاتنا التكشيفية (جدول ٦،٣)

| Relationships discussed | Word associations |
|-------------------------|-------------------|
| Word forms | Word derivatives |
| synonyms | similar |
| antonyms | contrast |
| hierarchical | superordinate |
| | subordinate |
| coordinate | coordinate |
| whole-part | whole-part |
| cause and effect | cause and effect |
| instruments | verb-object |
| materials | material |
| similarity | similarity |
| genetic | — |
| — | assonance |

الحاجة إلى تعرف العلاقات الدلالية

رأينا في بداية هذا الفصل أن الحاجة إلى أفراد ذاتية العلاقات الدلالية يمكن أن تظهر، ولكن في ضوء التحليل المفصل الذي أجريناه الآن، يجدر بنا أن نعيد صياغة المشكلة، من وجهتين مختلفتين للنظر. لقد بدأنا من وجهة نظر الباحث محاولين إجراء بحث في نظام مبنى على الحاسب، ومستخدمين نصوص الوثائق، أو أجزاء من نصوص الوثائق التي تشتمل عليها مجموعتنا. والسؤال هو: أى مجموعة من المصطلحات يجب أن نستعملها لكي نضمن أننا قد غطينا كل المعالجات (المداخل) الممكنة لمفهوم ما؟ يمكن أن نصوغ السؤال بطريقة أخرى. لو أن المصطلح الذي فكرنا فيه أولاً لم يسترجع الوثائق التي نريدها (أوربما لم يسترجع أية وثائق على الاطلاق!) ما المصطلحات الأخرى التي يمكن أن نستخدمها بدلا منه؟ من الواضح أنه سيكون مفيدا في هذا الوضع أن يكون لدينا قائمة تضم المصطلحات وتعرض العلاقات بينها، لكي تقترح علينا المصطلحات البديلة.

قائمة المصطلحات هذه - لغتنا التشفيرية - هي ذات قيمة مساوية بالنسبة للمكشف الذي يحاول تكويد وثيقة ما بطريقة تمكن الباحثين الذين يعتبرونها مفيدة من الوصول إليها. وقد يكون المفهوم الذي يحاول المكشف تكويده غير مألوف، وفي هذه الحالة سيكون من المفيد أن نكون قادرين على شق طريقنا خلال شبكة من الموضوعات المترابطة إلى أن نجد المصطلح الأنسب. وسوف يُدَلُّ كذلك على المصطلحات التي يرجح أن نكون قد استخدمناها في تكويد وثائق سابقة تعالج الموضوع نفسه.

وجدير بالملاحظة أنه إذا كان المفهوم الذي نعالجه له اسم معرف بشكل واضح يستخدمه المؤلفون ويعرفه الباحث، فإن لا يكون هناك حاجة إلى لغة تكشف بما تحتوى عليه من شبكة علاقات؛ وسوف نحسن صنعا إذا استخدمنا الحاسب لمضاهاة مصطلح البحث مع التسجيلات في قاعدة البيانات. ومع

ذلك، فليست هذه هي الحالة دائما، وسوف نحتاج غالبا إلى المساعدة في إجراء البحث، وبصفة خاصة إذا كنا نحتاج إلى استدعاء مرتفع. وسوف نجد أيضا أن هناك حاجة إلى لغات تكشف للأغراض الأخرى، مثال ذلك: في ترتيب الكتب على رفوف المكتبة، أو في إنتاج كشاف مطبوع. ولذلك فيجب أن نستمر في دراسة طرق ابتكار لغات التكشيف.

إبراز العلاقات الدلالية

رأينا حتى الآن أن لغتنا التكشيفية ينبغي أن تتألف من قائمة من المصطلحات تدل على المفاهيم التي نرغب في تضمينها، مع مجموعة من العلاقات التي تربط المجموعات المختلفة من المصطلحات. وعلينا الآن أن نتناول مشكلة: كيف نرتب مصطلحاتنا، وكيف نبرز العلاقات التي توجد بينها.

أبسط الطرق لترتيب مصطلحاتنا هي الترتيب الهجائي، كما في المعجم. ومع ذلك، فمن الواضح أن الترتيب الهجائي لا يمكن أن يبرز أي نوع من العلاقة فيما عدا العلاقة العرضية الخاصة بالأتان معا بالكلمات التي لها الجذر نفسه، وهذا لا يعالج على أحسن الفروض إلا جزءا صغيرا من المشكلة. والإجابة هي أن نضيف سلسلة من الروابط، يطلق عليها عادة الإحالات التبادلية-Cross-refer-ences، التي تفيد في إبراز العلاقات الدلالية أمام انتباه المستفيد.

تعني علاقات التقابل ضمنا اختيار شكل واحد على أنه المصطلح المفضل، كما سبق أن رأينا، ولهذا نعد إحالة تبادلية تحيل من المصطلح غير المفضل إلى المصطلح المفضل.

Footpaths

USE Trails

Book - marks

USE Bookmarks

وهذه الأمثلة مأخوذة من قرمك، والأول منها يحيلنا إلى مصطلح مفضل.

والثانى هو هجاء مفضل، وقد يبدو تافها إلى أن نفهم أن الحاسب عند ترتيب المصطلحات يعامل الشرطة كمقابل للمسافة، ومن ثم توجد: Book - marks بين: Book margins و Book numbers، قبل الرأس المعنى بحوالى ١٥٠ رأساً فى القائمة! وفى الفهرس المبني على الحاسب، من الأيسر تماماً بالنسبة للمستفيد الذى يذهب فى بحثه إلى المصطلح غير المفضل أن يُنقل بطريقة آلية إلى الرأس المفضل. وفى الفهرس البطاقى أو الفهرس المطبوع يجب على المستفيد أن يقوم بالانتقال بنفسه ودون مساعدة.

ونحن فى لغة الكشاف، نحتاج إلى إبراز عكس هذه الاتجاهات لكى نساعد المكشف، مرة أخرى باستخدام العرف أو التقليد (الأسلوب) UF، أى استخدم ل Use For :

Trails

UF Bridle paths

Foot tuails

Footpaths

Horse trails

Paths

Bookmarks

UF Book - marks

Paths

فى المثال الأول، تعامل Bridle paths، و Foot trails، و Horse trails، تعامل على أنها مقابل (بديل) لـ Trails، رغم أنه قد يظن أنها أكثر تخصيصاً. وهذه مسألة منفعة عملية؛ فإذا لم يكن هناك إنتاج فكرى كثير عن رأس مخصص، فإنه يمكن أن يوجد تحت الرأس الأوسع بصورة مناسبة ومريحة إلى حد معقول دون أن نزيد من عدد الرؤوس زيادة غير ملائمة. هذه الاتجاهات العكسية لاتظهر عادة فى الفهرس، ولكن هذا أمر يتعلق بممارسة تاريخية أكثر منه بالاعتبارات النظرية.

العلاقات الرتببة وعلاقات القرابة/ النسب

كما رأينا، فإن علاقات التقابل هى فى اتجاه واحد أساسا، تأخذنا من المصطلحات غير المفضلة إلى المفضلة. وعلى عكس ذلك، فإن النوعين الآخرين من العلاقة لهما اتجاهان، وهذا يعنى معالجة مختلفة بعض الشيء. سوف نحتاج إلى الاحالات فى الاتجاهين معا. ولتأخذ بعض الأمثلة عن العلاقات الرتببة، إذا كان لدينا مواد فى فهرسنا عن الشمس والنظام الشمسى فسوف نحتاج إلى جذب انتباه القارئ إلى هذا:

Solar system

| | |
|----|-----------------------------|
| BT | Milky Way |
| SA | names of individual planets |
| NT | Asteroids |
| | Comets |
| | ... |
| | Planets |
| | ... |
| | Sun |

فإذا رجعنا إلى Planets فسوف نجد:

Planets

BT Solar system

NT Earth

حيث م أ BT اختصار: المصطلح الأوسع، و م ق NT: المصطلح الأضيق. ويمكننا أن نرى أن العلاقات تبادلية: فكل م أ له م ق مرتبط به. ويمكن أن نرى كذلك أن ثمة أمثلة على العلاقات الأصلية (الشاملة) والعلاقات التجزئية أو التبعية؛ فالأرض هى عضو فى القسم الأجرام planets، ولكن الكويكبات Asteroids، والمذنبات Comets، والأجرام Planets، والشمس، هى كلها أجزاء من طريق اللبن Milky Way. وتستخدم ن أ SA [أنظر أيضا See also] فى قرمك للإحالات التى يمكن أن تصبح غامرة أو ساحقة:

Flowers

SA names of flowers, e.g. Carnations; Roses; Violets.

مثل هذه الاحالات الشاملة هي غير مرضية إلى حد ما، من حيث أنها تتركنا غير واثقين أى رؤوس سوف نبحسها بنجاح؛ ومن ناحية أخرى، فينبغى ألا نتوقع من لغة التكشيف أن تؤدي عمل دائرة معارف. وفي مجموعة صغيرة، يمكن أن نتجنب الاحالات الشاملة كلية، حيث أن عدد الرؤوس التي سوف يحال إليها سوف يكون محدودا.

فإذا أخذنا مثالا آخر، فإننا نجد:

Pressure Vessels

UF Containers, Pressurized

Pressurized containers

BT High pressure (Science)

Pressure

RT Autoclaves

Boilers

Caissons

Compressed air

Steam - boilers

Tanks

NT Gas cylinders

Hyperbaric chambers

Nuclear pressure Vessels

وهنا نجد أنه بالإضافة إلى استعمال USE، م BT، و م ق NT لدينا م RT المصطلح المترابط. وهذان الاثنان هما علاقتان متبادلتان، وكل م م يقابله م م

في القائمة كلها. في طبعات قرمك السابقة على الطبعة ١١، في ١٩٨٨، كان يستخدم نظام آخر مختلف، ولا يزال يوجد في بعض الفهارس المطبوعة أو البطاقية وبدلا من استعمال USE نجد انظر See، ونجد X تمثل استخدام من Use From. واستخدمت إحالة انظر أيضا See also بدلا من: م أ، م م، م ق، ولكن العكس كان XX. وهكذا يكون لدينا:

Forging

See also Blacksmithing

XX Blacksmithing

Blacksmithing

See also Forging

XX Forging

وقد أدخل النظام القديم في الطبعة ٥ في ١٩٤٨، ولكنه كان مربكا للمستفيدين. والنظام الحالي، الذي استخدم بشكل كامل لأول مرة في:

the EJC Thesaurus of engineering terms

في سنة ١٩٦٤ (الفصل ٢٥)، هو تحسين كبير؛ وإن تبنى قرمك له يعني أن استخدامه، الذي أوصت به كل المعايير القياسية المختلفة لبناء المكانز، قد أصبح الآن عالميا حقا.

مقارنة بين الروابط في لغات التكشيف المختلفة

قد يكون مقيدا وتعليميا في الوقت نفسه أن نقارن شبكات الإحالات التبادلية في اثنتين من لغات التكشيف، وقد اقترح كوشان Kochen وتاجليا كوزو Tagliacozzo مقياسين لهذا الغرض^(١٨). والأول منهما هو نسبة الارتباط Connectedness ratio:

عدد المصطلحات المربوطة بالمصطلحات الأخرى كجزء من العدد الإجمالي للمصطلحات في لغة التكشيف.

والثانى هو إمكانية الوصول accessibility :

متوسط عدد المصطلحات التى تقود إلى مصطلح معين.

وترتبط إمكانية الوصول باتساع شبكة الإحالات، فى حين أن نسبة الارتباط هى مقياس لعدد الأيتام Orphans، وهو مصطلح صُكَّ للدلالة على المصطلحات التى ليس لها روابط^(١٩). مثل هذه المصطلحات يمكن أن توجد فقط، بطبيعة الحال، إذا نحن فكرنا فيها لأنفسنا، حيث أن أحدا من المصطلحات الأخرى التى نفكر فيها لن يقودنا إليها من خلال التعريف. ومن بين ٣٦٠٠٠ مدخلا فى الطبعة ٧ من قرمك، هناك حوالى ٧٠٠٠ من الأيتام، وهى تبدو نسبة عالية؛ على أى حال، لقد وجد أوستن Austin نسبة مشابهة إلى حدما فى بريسيس، وهو نظام أكثر إحكاما وانضباطا بكثير من قرمك.

لغة المداخل ولغة الكشف

لا بد أن يكون قد اتضح من المناقشة السابقة أن المصطلحات التى ترتبط بواسطة علاقات التقابل تختلف عن النوعين الآخرين، من حيث أننا نختار مصطلحا واحدا مفضلا ونستخدم هذا المصطلح فقط فى كشافنا، فى حين أن النوعين الآخرين من العلاقة يردان بين مصطلحات تستخدم كلها فى الكشف. مثال ذلك: لو أننا اخترنا Pressure Vessels؛ أوانى الضغط على أنها مصطلحنا المفضل، وليس: Pressurized Containers، فإن كل المعلومات التى توجد فى مجموعتنا عن ذلك الموضوع سوف تُكشَف بواسطة المصطلح Pressure Vessels وليس بالمصطلح Pressurized Containers؛ والمستفيدون الذين يفكرون فى الأخير سوف يجدون التوجيه USE Pressure Vessels. وبالمقابل، فإننا سوف نستخدم Compressed air أو Autocalves باعتبارهما مصطلحين مناسبين، حيث أنهما مربوطان مع Pressure Vessels بواسطة إحالات المصطلح المترابط. ولذلك فإننا فى لغتنا التكوينية سوف يكون لدينا كل من المصطلحات المفضلة، التى تستخدم للتكشيف، والمصطلحات غير المفضلة، التى لاتستخدم. وتكون

المصطلحات المفضلة بذاتها لغة الكشاف index Vocabulary، في حين أن المصطلحات المفضلة والمصطلحات غير المفضلة يكونان معا لغة المداخل entry vocabulary. وإن لغة المداخل مهمة جدا، وهناك مناسبات كثيرة سوف نقرر فيها لسبب أو لآخر ألا نستخدم مصطلحا بالذات، بل أن نستخدم بدلا منه مصطلحا موجودا بالفعل في لغة الكشاف عندنا. ومن الجوهري أن يكون كل من المكشفين والباحثين قادرين على الوصول بسهولة إلى لغة الكشاف دونما حاجة إلى معرفة المصطلحات المفضلة سلفا. وقد بين مشروع مدلاز للتقويم^(٢٠) بوضوح شديد أن أية لغة مداخل غير كافية تقود إلى إخفاقات في كل من الاستدعاء والصلة، وأكد لانكستر Lancaster^(٢١) نقطة أن المصطلحات في لغة المدخل يجب أن تعكس كلا من السند الأدبي-literary war rant والسند الاستفساري euquiry warrant؛ وبمعنى آخر تلك المصطلحات التي نجدها في الانتاج الفكرى وكذلك أيضا المصطلحات التي يستخدمها القراء الباحثون عن المعلومات. يجب أن نكون على علم بالمصطلحات التي يستخدمها المستفيدون بنظامنا لا سترجاع المعلومات كما تلك المصطلحات التي يستخدمها المؤلفون الذين تكشف أعمالهم. وإن الافتقار إلى لغة مداخل مناسبة يعكسها الفشل في محاولة الوصول إلى معلومات في كتاب ما أعد له كشاف ضعيف أو لم يعد له كشاف على الاطلاق. وإن سبب إخفاقاتنا في الاستدعاء واضح جدا في هذا الوضع على الأقل!.

بناء المكنز

يتألف أبسط أنوا لغة التكشيف من قائمة من المصطلحات التي تعبر عن مفاهيم مفردة، وتبين العلاقات الدلالية بينها. ويطلق على مثل هذه القائمة عادة المكنز Thesaurus^(٢٢)، رغم أننا سوف نرى أثناء دراسة بعض الأمثلة النموذجية أنها لا تلتزم بصرامة بفكرة المفاهيم المفردة. وقد يكون لهذا قيمة عملية. مثال ذلك: نجد في مكنز إريك ERIC «فصول الدراسة الالكترونية» كراس مسبق؛ وهذا يتجنب النقطة الزائفة: التجهيزات الإلكترونية في فصول الدراسة المتنقلة؛

وقد ناقشنا في هذا الفصل النقاط الأساسية التي يجب مراعاتها في بناء مثل هذه القائمة، ويمكننا أن نرى الآن كيف تطبق هذه النقاط. مستخدمين أساليب التحليل نفسها التي استخدمناها سابقا. ولنشرع الآن في دراسة العناوين الواردة في قائمة ٦,٤. سوف يظهر بسرعة أنها تقع في جماعتين رئيسيتين، الأدب وعلم المعادن، وأن كلا من هاتين الجماعتين متجانسة ومتمايزة؛ مثل هذه الموضوعات المتجانسة والتممايزة يطلق عليها الأقسام الأساسية basic classes. والخطوة التالية هي تناول كل قسم أساسى على حدة للنظر في الأسس التي يمكن استخدامها لتحليله فيما وراء ذلك.

فإذا نظرنا إلى العناوين بعناية، فسوف نجد في الأدب مصطلحات مثل: الألماني، والفرنسى، والإنجليزى، والأسباني واللاتيني، وأنها تُكوّن جماعة من اللغات؛ وأن المسرحية، والشعر، والروايات تقترح جماعة طابعها الشكل الأدبي Literary Form؛ في حين أن القرن السابع عشر، و ١٩٠١ - ١٩١٥ و ١٨٠٠ - هي عصور زمنية وإذا استعرضنا الجماعة كلها، فسوف نجد أن كل المفاهيم المهمة تقع في واحد أو آخر من هذه الجماعات (الفرعية). وفي علم المعادن، نجد عددا من المصطلحات التي تدل على المعادن، سواء كانت أفرادا أم عائلات (مثال ذلك غير الحديدية)، والمصطلحات الأخرى التي تدل على أن نوعا ما من العمليات ينفذ؛ مثل المعالجة بالحرارة heat treatment. هذه الجماعات، التي نتجت عن تجميع المفاهيم التي ترتبط بالقسم الأساسى بالطريقة نفسها يطلق عليها أوجه Facets الموضوع، وتعرف المفاهيم المفردة في نطاقها بالبؤرات Foci (جمع بؤرة).

فإذا ما حللنا الموضوعات إلى الأقسام الأساسية وإلى الأوجه المناسبة، فيمكننا تحديد ذاتية العلاقات في نطاقه الأوجه. ومن المحتمل أن أى وجه سوف يحتوى على أمثلة علاقات التقابل: كلمتان أو أكثر يمثلان المفهوم نفسه. إن عملية تحديد ما الذى تمثله الكلمات هي التي يتم بموجبها تحديد ذاتية المترادفات وأشباه المترادفات، وتمكننا من أن نقرر ما العمل الذى يجب أن

نقوم به - نختار أحد المصطلحات كمصطلح منفصل ونستخدم المصطلحات الأخرى فقط في لغة المداخل، أو نحاول أن نعرفها بطريقة تجعل استخدامها يدل ولو بصورة محدودة على مختلف المفاهيم. مثال ذلك: في العناوين الواردة في جدول ٦,٤ نجد «القصة» و«الرواية»؛ و«المسرحية» و«كتابة المسرحية»؛ «الشعر» و«النثر» «المعالجة الحرارية» و«المعالجة بالحرارة». ويمكننا أيضا أن نحدد ذاتية علاقات رتيبة: «المعادن غير الحديدية» و«الألومنيوم»، «النحاس»، «بريليوم»؛ «الحديد والصلب» الذي يمكن تحديد ذاتيته على أنه من «المعادن الحديدية Ferrous»؛ «المعادن النادرة» التي هي جماعة ثالثة؛ «المعالجة بالحرارة» هي طريقة «للصناعة»؛ وقد رأينا من قبل أن «الألمانية» و«الفرنسية» و«الانجليزية»، و«الأسبانية» و«اللاتينية» تقع كلها ضمن الرأس العام: اللغات. ومن العلاقات الرتيبة يمكننا أن نحدد ذاتية بعض علاقات القرابة؛ مثال ذلك: كل واحد من المعادن غير الحديدية له علاقة متساوية في الرتبة مع المعادن الأخرى. وهكذا يكون لدينا معلومات كافية تجعل من الممكن أن نبدأ بناء مكنز يغطي المجال الموضوعي لقسم أساسى. ومن الواضح أن لدينا قدرا كبيرا من العمل يجب أن نؤديه قبل أن نقول إن المكتز قد أصبح كاملا أو حتى يمكن استخدامه، ولكن إذا ماتم لنا تعرف البنية الأساسية للأوجه وسلاسل الرتب فى داخل هذه الأوجه، فلن يكون من الصعب إضافة المصطلحات الجديدة حينما تظهر. ويجب أن نبني شبكتنا من الروابط، والمبنية على سلاسل الرتب، ويفضل أن نتجنب روابط القرابة أو النسب التي لا تستمر ضمن هذا الاطار.

شكل ٦,٤ التحليل إلى الأقسام والأوجه الأساسية

-
- 1 The study of literature
 - 2 Select methods of metallurgical analysis
 - 3 The growth of the English novel
 - 4 Elements of heat treatment of metals
 - 5 Modern drama 1800-
 - 6 A textbook of metallurgy
 - 7 Playwriting
 - 8 Methods for the analysis of aluminium
 - 9 The poet's task
 - 10 Iron and steel
 - 11 The French drama of today
 - 12 The metallurgy of beryllium
 - 13 The literature of the Spanish people
 - 14 The manufacture of iron and steel
 - 15 A history of English drama 1600-1900
 - 16 Equipment for the thermal treatment of non-ferrous metals
 - 17 English literature of the twentieth century
 - 18 Rare metals handbook
 - 19 The background of modern English poetry, 1901-1915
 - 20 Methods for the analysis of raw copper
 - 21 Latin literature
 - 22 Heat treatment of aluminium
 - 23 English literature and its readers
 - 24 Heat treatment of steel
 - 25 Some principles of fiction
 - 26 Copper: the science and technology of the metal
 - 27 A short history of German literature
 - 28 Methods for the analysis of iron and steel
 - 29 The temper of the seventeenth century in German literature
 - 30 Twentieth century German verse
 - 31 A few facts about aluminium
 - 32 The decline of the Spanish novel, 1516-1600
-

ومن علاقات التقابل التي نحدد ذاتيتها يمكننا أن نقرر أي روابط استعمل USE أو استعمل من UF سوف نعد. مثال ذلك:

Heat treatment

UF Thermal treatment

Thermal treatment

USE Heat treatment

ويجب أن تتدرج الروابط الرتبية، أى: تربط خطوة واحدة فى وقت واحد. مثال ذلك: حددنا ذاتية ثلاث جماعات كبيرة فى نطاق وجه المعادن: المعادن الحديدية، والمعادن غير الحديدية والمعادن النادرة. ولذلك فسوف يكون لدينا:

Metals

- NT Ferrous metals
- Non - ferrous metals
- Rare metals

فى داخل المعادن الحديدية سوف يكون عندنا الحديد والصلب، وفى داخل المعادن غير الحديدية عندنا الألومنيوم، والنحاس، والبريليوم. وحتى الآن لانجد أى مثال مخصص على المعادن النادرة. وهكذا فإننا ينبغى أن نجد الروابط الآتية:

Ferrous metals

- NT Iron
- Steel

Non - ferrous metals

- NT Aluminium
- Beryllium
- Copper

وما يجب علينا تجنبه هو روابط مثل:

Metals

- NT Aluminium

فهذه تحذف خطوة التقسيم، وتفتح الطريق إلى وضع يكون فيه كل مصطلح مرتبطاً بالآخر - فى هذه الحالة، طبعا، يتساوى الوضع كما لو كنا بدون روابط على الاطلاق. (إذا لم يكن منطق هذه واضحا بصورة مباشرة، فيجدر أن نعطي

شيئا من التفكير للمسألة؛ وسوف ندرسها بعد مرة ثانية مرتبطة ببعض المكانز التي سوف نصفها).

ويمكن أن نتعرف الروابط المتساوية في الرتبة باعتبارها مصطلحات مترابطة:

Aluminium

RT Beryllium

Copper [and the same treatment for Beryllium and Copper]

Ferrous metals

RT Non - ferrous metals

Rare metals

ومع ذلك فيجب ألا يكون عندنا:

Aluminium

RT Iron

لأن الحديد في سلم مراتب آخر.

وقد نُصِحَ أيضا بتوخى الحذر في إعداد روابط م م بين الرؤوس المتساوية في الرتبة. إليك المثال الآتي:

Vehicles

NT Boots

Ships

في الاستعمال العام يوجد تداخل بين الرأسين «السفن» و«القوارب»، ولذلك فينبغي أن نُذكَرَ المستفيدين الذين يجدون واحدا منهما بأن يذهبوا إلى الثاني أيضا. كذلك فإن اشجار البلوط "oaks" وأشجار الدردار "elms" هما مصطلحان أضيّق من «الأشجار»، ولكن ليس ثمة تداخل في المعنى، ولذلك فليس هناك

ضرورة لأن نُذَكِّرُ المستفيدين الذين يجدون واحدا بوجود الآخر. وهذا صحيح تماما، ومع ذلك فإنه يعنى أننا قد يكون علينا أن نصدر حكما حول تضمين رابطة م م بين المصطلحات وذلك بالنسبة لكل مصطلح فى سلم الرتب. والطريقة المقترحة أنفا هى وسيلة آلية لتوليد روابط م م ولا تتطلب ممارسة الحكم أو التقدير؛ ومع ذلك فهى تمكننا من أن نسيطر على شبكتنا من الروابط، وتضمن أنها كاملة، دون إحالات تبادلية غير ضرورية.

ولكى نكمل المكنز، فيجب أن نرتب مصطلحاتنا فى ترتيب هجائى. وقد نضيف أيضا سياقات إضافية، ولكن العمل الأساسى فى بناء المكنز يكون قد تم. وقد تتضمن السياقات الإضافية كشافا دائريا Rotated، لكى يكشف عن الكلمات المحتجبة فى المصطلحات المتعددة الكلمات، كما يتطلب عرضا رتبيا، لكى يبين العلاقات التصنيفية. وفى حين أن هاتين الإضافتين ليستا جزءا أساسيا من المكنز، فإنهما فى معظم الحالات يكون لهما فائدة كبيرة، كما سوف يتضح من الأمثلة العملية التى ناقشناها فى الفصلين ٢٥ و٢٦. وبطبيعة الحال يجب أن نتذكر أن المناقشة السابقة ليست إلا توضيحا فقط؛ ففى الأدب يوجد أكثر من ثلاثة أوجه، وفى علم المعادن أكثر من وجهين، وهذا يمكن أن يستبين من مزيد من تفحص الانتاج الفكرى. ومع ذلك فإن الطريقة تبقى هى هى، وتزودنا بطريقة صحيحة وبسيطة إلى حد كبير لتطوير لغة التكشيف. ويوجد الآن برامج حاسب مختلفة للمساعدة فى تطوير وصيانة المكانز، وهى تتفاوت بين برامج أساسية وبرامج عالية التطور التى يحتاج إليها لصيانة مكنز متشابك؛ وقد ناقش ميلستد Milstead^(٢٣) متطلبات هذه البرامج بالتفصيل.

قد يقدم الانتاج الفكرى للعلوم الاجتماعية مشكلات خاصة إذا حاولنا أن ننفذ هذا النوع من التحليل. فالكلمات غالبا ما تؤخذ من مجالات موضوعية أخرى وتستخدم بمعان خاصة - لغة jargon - بحيث يكون استخدامها بتلك

الطريقة محدودة اربما بفرع واحد من العلوم الاجتماعية. الصحة السياسية Political Correctness، تفرض أنه يجب ألا تستخدم كلمات معينة؛ مثال ذلك: الأمية illiteracy قد يظن أنها تعنى ضمنا الطعن أو القبح، ولذلك يستخدم بدلا منها غير القارئین non-readers، الأمر الذى يعنى أن مصطلحا واضح المعنى بالنسبة لمعظم الناس قد حل محل مصطلح غامض، ويمكن أن يشير إلى أكثر من حالة واحدة. وقد يكون من الضروري أن نعد حواشى تعريفات definition notes DF وكذلك حواشى مجال Scope notes SN، بحيث يصبح المكنز حينئذ توصيفيا لا وصفيا. وكثير من الإنتاج الفكرى فى العلوم الاجتماعية يقع فى الفئة التى تسمى «الانتاج الفكرى الرمادى» ولا تغطيها المصادر البيولوجرافية القياسية؛ ولو استخدم المكنز بواسطة العاملين فى المجال الميدانى الذين ليست لديهم المهارات الضرورية لتكشيف المادة، فإن النتيجة سوف تكون أقل من مرضية^(٢٤).

وتبدأ الأساليب السابقة بدراسة للكلمات المستخدمة فى مجال موضوعى معين، ولكنها تتحرك تجاه فكرة تكشيف المفهوم وذلك عن طريق حذف المترادفات وتعرف العلاقات. فإذا التزمنا ذلك، فإننا يمكن أن نحدد ذاتية المفهوم الأساسى نفسه حيثما ورد، حتى وإن لم يكن هناك أى عامل مشترك فى الكلمات المستخدمة له. ويمكن أن نربط: Stream, Current, Flux, Flow, evolution، ونربط: tion باعتبارها مظاهر للحركة motion؛ ونربط:

expurgation, disinfection, refining, Bowdlerization, white washing

باعتبارها مظاهر للتنظيف Cleaning. مثل هذه القائمة هى المعنى الذى قصده روجيه بالكلمة مكنز thesaurus فى كتابه:

Thesaurus of English words and phrases

والذى أخذنا منه المثالين السابقين؛ وهناك أمثلة أخرى^(٢٥).

الترتيب المقنن .

إن المناقشة السابقة عن ابراز العلاقات الدلالية ترتبط بطريقة واحدة من طرق الترتيب: الترتيب الهجائي . ويمكننا أيضا أن نبرز العلاقات بواسطة: التجاور المكاني juxtaposition، أى تجميع المفاهيم المترابطة فى ترتيب مقنن Systematic لتكوين خطة تصنيف Classification Scheme . مثل هذا الترتيب سوف يبرز العلاقات الرتبية وكذلك علاقات التساوى فى الرتبة، وقد يبرز علاقات أخرى كذلك مثل الآلات والمواد . وبهذه الطريقة، فإن جانبا جوهريا من بنية الاحالات التبادلية المطلوبة للترتيب الهجائي - سوف يحذف، لأن العلاقات معروضة أصلا بالطريقة التى تجمع بها المفاهيم . ومن المبعثاد أننا نرتب الكتب على الرفوف بهذه الطريقة لكى نساعد المستفيدين، الذين فيما نأمل سوف يجدون كل الكتب التى يهتمون بها على الرفوف معا فى المكان نفسه .

ومع ذلك فإن هناك ثمنا يجب أن ندفعه لهذه الميزة . فإذا جمعنا مصطلحاتنا المفضلة بطريقة مقننة، فإن الترتيب الذى تكون عليه لن يكون واضحا بذاته، وسوف نكون مضطرين إلى إضافة رمز notation أو لغة كودية Code Vocabulary لكى نوضح أو نعرض الترتيب، ولكى يمكننا من الوصول إلى المفاهيم وسط الترتيب المقنن . وتصبح لغة المداخل الآن مضاعفة الأهمية، لأنها لانتحتوى على كل المصطلحات غير المفضلة فحسب، ولكن لأنها مرتبة هجائيا، فهى تشكل وسيلتنا الوحيدة للوصول إلى الترتيب المقنن، عن طريق اللغة الكودية . ونحن بحاجة إلى أن نصل إلى المصطلحات التى نهتم بها فى لغة المداخل، والتى سوف تدلنا على الكودات التى استخدمت للدلالة عليها:

| | | |
|--------------------|--------------|-------|
| Electronics | 621. 381 | (DDC) |
| Cyclotrons | 621. 384. 61 | (UDC) |
| Preaching | PXP | (BC1) |
| Disease (Medicine) | L:491 | (CC6) |
| Amplifiers | TK 6565. A55 | (LCC) |

ويتم التعبير عن علاقات التقابل ببساطة عن طريق بيان الكود نفسه لكل؛ وفي الحقيقة فإن كل المداخل في لغة المداخل يمكن اعتبارها علاقات تقابل، من حيث أنها توضح الرأس الذي استخدم للترتيب (في هذه الحالة قطعة من الرمز) بالنسبة لكل من المصطلحات المفضلة وغير المفضلة. في جداول الخطة، أى قائمة مصطلحات لغة الكشاف في ترتيب مقنن، سوف نجد المصطلحات المفضلة فقط.

وتظهر مشكلة أخرى لأن المفهوم، كما رأينا من قبل، يمكن أن يظهر فى أكثر من سلم رتب واحد، حيث نجد كلا من العلاقات الأصيلة (الشاملة)، والدائمة، وشبه الشاملة أو الأصيلة التى تمثل التطبيقات. وعلى هذا فإن المفهوم الأساسى نفسه قد يمثله أكثر من كود واحد، وهذا يعتمد على السياق الذى يظهر فى داخله:

Tobacco

| | |
|----------------|--------|
| botany | 583.79 |
| hygiene | 613.8 |
| social customs | 394.1 |

والترتيب المقنن يمكنه أن يعرض عددا من فئات العلاقات التى حددنا ذاتيتها، إما بواسطة التجاور المكانى فى الجداول أو بواسطة التجاور المكانى المكمل بمدخل فى التسلسل الهجائى للغة المداخل. ومع ذلك، فإن هذه لاتغطى كل جماعة القرابة/ النسب، والتى قد يكون بعضها مخبأ حقيقة بواسطة الترتيب. والطريقة الوحيدة التى يمكن بواسطتها وضع هذه أمام انتباه المكشف أو الباحث هى من خلال الإحالات التبادلية فى الجداول أو فى لغة المداخل. ولسوء الحظ، فإن مثل هذه الاحالات هى استثناء وليست قاعدة فى معظم خطط التصنيف؛ وقد يكون هذا انعكاسا لحقيقة أننا لم نبدأ إلا فى السنوات الأخيرة فقط فى بيان حقيقة هذه العلاقات التى يمكن أن تحدث بين المفاهيم.

وظيفة العلاقات الدلالية

كما سبق وأن رأينا فإن اشتغال نظامنا للتكشيف على شبكة من العلاقات الدلالية يمكننا من تحسين صياغتنا للتكشيف أو البحث عن طريق اقتراح مصطلحات بديلة أو إضافية يمكننا استخدامها. في مرحلة التكشيف، نحاول أن نعرف سلفا الطرق التي سوف يرغب المستفيدون لاحقا في أن يجدوا من خلالها الوثائق؛ وفي مرحلة البحث، نحاول أن نحقق مضاهاة أفضل بين صياغتنا للسؤال والمصطلحات التي يستخدمها المكشفون أو المؤلفون.

ولذلك فإن العلاقات الدلالية تكون وظيفتها الأولى تحسين الاستدعاء، ولكن قد يكون هذا على حساب الصلة. فإذا كان علينا أن نستخدم مصطلحا آخر بدلا من المصطلح الذي فكرنا فيه في البداية، فقد يكون بديلا أدنى منزلة، وإن لم يكن ذلك حتميا؛ قد يكون المصطلح الذي فكرنا فيه أولا هو نفسه الأفضل. ومن الصعب غالبا أن نجد على وجه الدقة المصطلح الصحيح أو المناسب للتعبير عن المفهوم، وبخاصة إذا ماتذكرنا أن الباحث يبدأ من وضع يكون فيه جاهلا نسبيا. ومع ذلك فإذا كنا متأكدين من المصطلح المناسب، فإنه حتى المرادف قد يكون أقل قبولا؛ ويمكن أن يحتاج بأنه في حالة المصطلحات التي لها معاني مفهومية Connotative فلن يكون هناك مرادفات كاملة أو مضبوطة. وإن إدماج علاقات المترادفات مع علاقات التقابل الأخرى، والتجميعات التي تحقق بواسطة العلاقات الرتبية وعلاقات القرابة/ النسب، هي في الأساس وسائل لضمان أن وثائق أكثر سوف تسترجع إجابة لسؤال، أي أنها وسائل لتحسين الاستدعاء. ومن ناحية أخرى، يمكن استخدامها كوسائل لتحسين الصلة عن طريق تمكين المكشفين أو الباحثين من الوصول إلى الاختيار الأمثل للمصطلحات التي تستخدم في وضع بعينه.

المراجع

- 1 Cleverdon, C. W., Mills, J. and Keen, E. M., *Factors determining the performance of indexing systems*. Cranfield, Aslib-Cranfield Research Project, 1966. 2v in 3.
- 2 *The concise Oxford dictionary*, Oxford University Press. 6th edn. 1976.
- 3 *Webster's seventh new collegiate dictionary*. Springfield, Mass. Merriam, 1971.
- 4 Berle, D., *The process of communication*, New York. Holt. Rinehart and Winston, 1960. 191-6.
- 5 Willetts, M., 'An investigation of the nature of the relation between terms in thesauri', *Journal of documentation*, 31 (3), 1975, 158-84.
- 6 Austin, D., *PRECIS: a manual of concept analysis and subject indexing*. 2nd edn. London, The British Library, 1984. 86-7.
- 7 Perry, J. W. and Kent, A., *Tools for machine literature searching: semantic code dictionary: equipment: procedures*. New York. Interscience Publishers Inc. 1958.
- 8 Lancaster, F. W., Elliken, C. and Connell, T. H., 'Subject analysis', *Annual review of information science and technology*, 24, 1989. 35-84.
- 9 Jones, K. P., 'Compound words in thesauri', *Journal of documentation*, 37 (2), 1981. 53-68.
- 10 Ref. 6 above, 62.
- 11 International Organization for Standardization. *Guidelines for the establishment and development of monolingual thesauri*, Geneva, ISO, 2nd edn. ISO 2788:1986.
- 12 British Standards Institution, *Guide to establishment and development of monolingual thesauri*, Milton Keynes, BSI, 1987. BS 5723:1987.
- 13 National Information Standards Organization. *American standard guidelines for the construction, format and management of monolingual thesauri*, Bethesda, MD, NISO, ANSI/NISO Z39.19: 1993.
- 14 Vickery, B. C. and Vickery, A., 'Online search interface design', *Journal of documentation*, 49 (2), 1993, 103-87.
- 15 Ref. 6 above. p80.
- 16 Coates, E. J., *Subject catalogues: headings and structure*, London, Library Association. 1960. (This has been reissued with a new preface, 1988.)
- 17 Miller, G. A., *Language and communication*, New York. McGraw Hill, 1951. Ch9.
- 18 Kochen, M. and Tagliacozzo, R., 'A study of cross-referencing', *Journal of documentation*, 24 (3), 1968. 173-91.
- 19 Sinkankas, G. M., *Study in the syndetic structure of the Library of Congress list of subject headings*, Pittsburgh University. Graduate School of Library and Information Science, 1972. (Pittsburgh studies in library and information sciences 2).
- 20 Lancaster, F. W., *Evaluation of the MEDLARS Demand Search Service*. Bethesda, MD. National Library of Medicine. 1968.

- 21 Lancaster, F. W., *Vocabulary control for information retrieval*, 2nd edn. Arlington, VA. Information Resources Press. 1986.
- 22 Aitchison, J. and Gilchrist, A., *Thesaurus construction: a practical manual*. 2nd edn. London, Aslib. 1987.
 Foskett, D. J., 'Thesaurus', in *Encyclopedia of library and information science* New York, NY., Dekker, v30. 1980. 416-63.
 Townley, H. M. and Gee, R. D., *Thesaurus making - grow your own word stock*. London. Deutsch, 1980.
- 23 Milstead, J. L., 'Thesaurus management software'. *Encyclopedia of library and information science*, 51. 1993. 389-407.
- 24 Hudon, Michèle, 'Term definitions in subject thesauri: the Canadian Literacy Thesaurus experience'. in *Classification research for knowledge representation and organization: proceedings of the 5th International study conference on classification research, Toronto, Canada, June 24-28 1991*, Williamson, N. J. and Hudon, M. (eds.), Elsevier, 1992. (FID 698). 255-62.
- 25 Neaman, J. and Silver, C., *In other words: a thesaurus of euphemisms*. London. Angus & Robertson, 1991. is an entertaining example.

التكشيف المعين ٢: التراكيب

في الأجزاء التي افتتحنا بها الفصل السابق رأينا أن البحث يتضمن نوعين من العلاقات: الدلالية وهي التي تنشأ من الحاجة إلى القدرة على البحث عن المصطلحات البديلة، والبنائية أو التركيبية الاعرابية التي تنشأ من الحاجة إلى القدرة على البحث عن نقاط تقاطع بين اثنين من الأقسام أو أكثر تعرفها المصطلحات التي تدل على مفاهيم محددة. وفي هذا الفصل سوف نهتم بصفة رئيسية بالطرق التي يمكن من خلالها أن ننفذ عملية التقاطع بين الأقسام أو ما يسمى الربط Coordination .

يمكننا معالجة المشكلة بطريقتين مختلفتين تماما. إذا استخدمنا المصطلحات من لغة تكشيف للدلالة على المفاهيم المفردة الموجودة، فإننا نكون قادرين على ربط المصطلحات في وقت البحث عبر استخدام الحاسب، كما رأينا في الفصل ٥. ويعرف هذا بتكشيف الربط اللاحق أو التالي post-co-ordinate indexing، وتوجد أصوله في الأشكال المادية مثل المصطلح الواحد Uniterm أو بطاقات المطابقة الضوئية التي حل الحاسب محلها الآن. قد نقرر، مع ذلك، أننا نريد أن نمثل الموضوعات التي تتألف من أكثر من مفهوم واحد كوحدة، عن طريق ربط المصطلحات المناسبة في وقت التكشيف؛ وتعرف هذه الطريقة بالتكشيف المسبق Pre-coordinate indexing، ولازال يستخدم على نطاق واسع جدا في الكشافات المطبوعة، وفي رؤوس الموضوعات مثل قرمك، وفي ترتيب الرفوف. وفي كلتا الحالتين فإن علينا أن نقرر إن كان علينا أن نبرز العلاقات التركيبية، وإذا كان الأمر كذلك، كيف نبرزها. وينطوي التكشيف المسبق (السابق) على مشكلات أكثر من اللاحق؛

وإذا استطعنا أن نرى كيف يمكن حلها بالنسبة للتكشيف السابق، فإن التكشيف اللاحق يصبح بسيطا نسبيا.

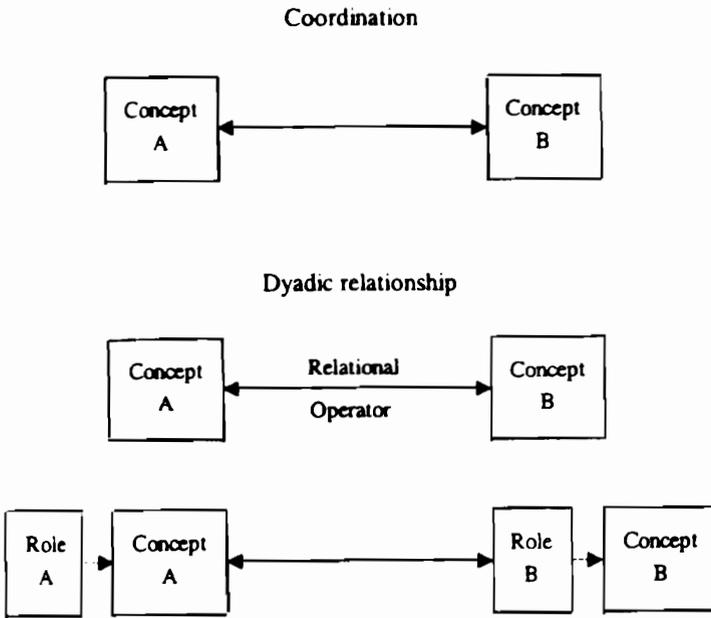
ومن الضروري أن نوضح مشكلة المصطلحات قبل أن نتقل أبعد من هذا فإن ما يطلق عليه الآن التكشيف اللاحق كان يعرف من قبل بالتكشيف المترابط Coordinate. ولم يتم معرفة أن كل أنواع التكشيف تنطوي على ربط للمفاهيم إلا بعد أن طورت الخلفية النظرية، وهي تتضمن الربط إما وقت التكشيف أو وقت البحث. وقد وضعت المصطلحات: المسبقة-Pre - coordi- rate، وهو الذى يستخدم للدلالة على التكشيف الذى تربط فيه المفاهيم وقت التكشيف، واللاحق، للدلالة على التكشيف الذى تربط فيه المفاهيم وقت البحث - وضعت هذه المصطلحات لتوضيح الفارق بينهما مع التأكيد على أنهما متشابهان من الناحية الجوهرية.

ويمكننا أن نبدأ بتناول بعض المشكلات العامة تماما والمتعلقة بالعلاقات بين المفاهيم. فى شكل ١، ٤ نرى أن كل العلاقات بين المفاهيم يمكن اعتبارها ثنائية أساسا: أى يمكن القول أنه حتى لو كان لدينا مفاهيم متعددة يحتوى عليها موضوع مركب Composite مثل:

the manufacture of multiwall Kraft paper sacks for the packaging of cement، يمكن أن ننظر إليه باعتباره سلسلة من أزواج المفاهيم ربطت معا. وهناك ثلاث طرق يمكن بواسطتها أن نعبر عن مثل هذه العلاقة. الأولى هى أن نقول ببساطة إن علاقة توجد، دون أن نخصص أى نوع من العلاقة يوجد: هذا هو نوع الربط الذى نحصل عليه من بحث بسيط بواسطة الحاسب عن ورود مصطلحين معا. فإذا استخدمنا رابطة و AND، فسوف نعلم أن الكلمتين تردان فى الوثيقة نفسها - وربما كانت علاقة ضعيفة جدا فى الواقع. والطريقة الثانية هى تعريف نوع العلاقة التى توجد بين مفهومين باستخدام معامل علاقة relational operator لربطهما. وقد ابتكر نظامان يستخدمان معاملات العلاقة، أحدهما وضعه ج. أ. ل. فرادان J. E. L. Farradane^(١) والثانى ج.

س. جاردن J. C. Gardin،^(٢) ولكن أيا منهما لم يترك أى تأثير مهم على الاتجاه العام لنظم استرجاع المعلومات، رغم أن بعض أفكارهما قد أثرت على عاملين آخرين.

شكل ١، ٧ - الربط - طرق بيان العلاقات الثنائية



ولقد بنى فرادان نظام على عملية التعليم وكيف تتطور. ويتألف نظامه من تسع معاملات، استخدمها لربط مصطلحات تكون analets. وثمة سمة مهمة هي أن analet لا يتعين أن يكون خيطاً طويلاً من المصطلحات؛ فهناك بنى ذات بعدين، وهي فكرة تبنتها فيما بعد بريسيس PRECIS. وقد زعم فرادان نفسه أن analets يسهل بناؤها، وهي تحتاج إلى دقيقتين؛ وقد أثبتت التجربة في مشروع ISILT^(٣) أن الرقم الحقيقي يجب أن يكون نصف ساعة!

وقد نشأت خطة جاردن من الحاجة إلى طريقة مشتركة للتكشيف لكي تستخدم في نطاق جماعة EURATOM. وقد أمكن تعرف أربعة أنواع فقط من العلاقات، وكان لابد أن تكون هذه الأنواع عامة جدا. وربما كان أكثر إسهامات جاردن فائدة هو تحديده لعلاقات السابقة a priori و اللاحقة a posteriori: تلك التي تنشأ من بنية المعرفة، وتلك التي تنشأ من الارتباط المؤقت للمصطلحات في وثيقة ما. وقد استخدمنا المصطلحين «الدالية» و«التركيبية» لوصف هذه العلاقات؛ وقد ذهب جاردن أبعد من ذلك حينما اقترح أن العلاقة نفسها يمكن أن تقع في أي قسم وهذا يعتمد على الوثيقة التي توجد فيها. والعلاقات الدالية باقية أو دائمة في العادة، ولكن العلاقات شبه الشاملة يمكن أن تقع في نطاق مازعمه جاردن.

دالات الدور

إن معاملات العلاقة ترتبط بالضرورة بالتكشيف المسبق، حيث أنها تربط بصورة دائمة المفاهيم المتضمنة. أما دالات الدور Role Indicators فهي مرتبطة بالمفاهيم التي تحيل إليها، ومن ثم يمكن استخدامها في التكشيف المسبق واللاحق. ومع ذلك فإنها توجد بصورة أكثر شيوعا في التكشيف المسبق، حيث يمكن أن تكون صريحة أو مباشرة، كما في قرمك، أو ضمنية، كما في تعدد، ولكن قبل أن ندخل في مناقشة عن كيفية استخدامها، يجب أن نعرف أولا ما الذي نعنيه بالكلمة دور role.

رأينا في الفصل السابع أنه كان من الممكن أن نحدد شخصية خمس فئات واسعة من المفاهيم، وكان هذا التحليل ذا قيمة من حيث أنه يعرض أنه بصرف النظر عن الصفات النعتية؛ فإن كل المفاهيم كان يعبر عنها بالأسماء. ومع ذلك، فإنه بالنسبة لمعظم الأغراض العملية لا يكون التحليل إلى هذه الفئات الخمس مفيدا إلا إلى نقطة ما، ثم نحتاج لتحديد شخصية أو ذاتية مزيد من الفئات المخصصة. ويمكن أن نفعل هذا على المستوى العام، لكي نعطي فئات ذات

تطبيق عام، مثل المادة الخام، المنتج؛ أو يمكن أن نبتكر فئات تقع في واحدة من هذه الجماعات العامة ولكن يمكن أن تعطى أسماء مخصصة لأنها ترتبط بمجالات موضوعية مخصصة، مثل المحاصيل (المنتج في المجال المخصص الزراعة). وعن طريق وضع المفهوم في واحدة من هذه الفئات، فإننا بهذا نعرف دوره بالنسبة للمفاهيم في الفئات الأخرى.

ويمكن أن نوضح هذه النقطة من خلال إعطاء بعض الأمثلة. «معالجة المعادن بالحرارة» يتضمن عملية operation هي المعالجة بالحرارة تتم على قابل patient بواسطة أداة أو وسيلة agent الحرارة؛ بستره اللين بدرجة حرارة فائقة، يعطينا الفئات نفسها بالضبط. «نوافذ الألومنيوم» تتضمن منتجاً product هو النوافذ مصنوعة من المادة الخام الألومنيوم: ويعطينا «أكياس البلاستيك» الفئات نفسها. ويمكننا أن نحلل «معالجة المعادن غير الحديدية بالحرارة»، «لحام welding أكياس البلاستيك»، «أفلام البولي إيثيلين» - يمكننا أن نحللها جميعاً بالطريقة نفسها بحيث تدخل في فئات تُعرَّف دورها بالنسبة للمفاهيم الأخرى المتضمنة. وإن مصطلحاً معيناً قد لا يدخل دائماً في الفئة نفسها، وهذا يعتمد على دوره في الموضوع الذي نرغب في تمثيله؛ مثال ذلك: في الموضوع «استخلاص الألومنيوم من البوكسيت» الألومنيوم هو المنتج النهائي end prod-uct والبوكسيت هو المادة الخام. والاستخلاص أو الاستنباط هو العملية.

وقد ابتكر مكشفون مختلفون أنظمة متعددة من الفئات، وسوف نناقشها بالتفصيل مع تناولنا لكل نظام. والنقطة التي تهمننا الآن هي أن هذه الفئات جميعاً تمثل أدواراً، وأنها تعرضها كما هي بطريقة صريحة مباشرة، عن طريق استخدام نوع ما من الرموز أو العلاقات، أو تعرضها بطريقة ضمنية، عن طريق موقعها النسبي. وفي النظم المسبقة قد نجد الطريقتين مستخدمتين؛ مثال ذلك: في الرأس المأخوذ من كبت BTI:

"BOTTLES; Polyethylene, blow moulded"

نجد أن الفاصلة المنقوطة قبل البولي إيثيلين تدل على أنه مادة، في حين أن الفاصلة قبل المشكلة بالنفخ تدل على أن هذا نوع من البولي إيثيلين، في حين أنه في رمز تصنيف ديوى العشري ٨٢٣ نعرف من دراسة بنية هذا الجزء من الخطة أن ٢ تمثل اللغة (الانجليزية) و٣ الشكل الأدبي في نطاق قسم الأدب الذي يمثله ٨. في المثال الأول نجد أن دالات الدور صريحة: وهي تظهر في شكل رموز إضافية. أما في الحالة الثانية، فإن دالات الدور ضمنية وتأتي في شكل رمز. وفي النظم اللاحقة، فإن دالات الدور، إذا استخدمت أصلاً، يجب أن تكون صريحة؛ مثال ذلك: قد نجد "Paint (B)" في كشاف يستخدم قائمة مصطلحات المعهد الأمريكى للبتروول، حيث تدل (B) على المنتج، أو «الاستخدام العلاجي للبنسيلينات» و«تحليل البنسيلينات» في الكشاف الطبى، حيث الأدوار تتهجى كما لو كانت رؤوساً فرعية.

نظام التوفيق

من الواضح أننا يمكن أن نستخدم دالات الدور مع المفاهيم المفردة، ورغم ذلك فإننا معنيون في العادة بالموضوعات المركبة، أى تلك التى تتضمن مفهوميين أو أكثر. ونحن نجمع هذه المفاهيم، فى النظم المسبقة، فى خيوط Strings من المصطلحات ترتبط بعبارات أو جمل فى اللغة الطبيعية. ولذلك فيجب أن يثار السؤال: لماذا لانستخدم عبارات من اللغة الطبيعية كما هى؟ لماذا نجشم أنفسنا مشقة بناء تراكيب مصطنعة كما لغة مصطنعة؟ (مصطنعة بمعنى أننا نتحكم فى شكل الكلمات، فنستخدم الأسماء أو الصفات فقط، ونختار المصطلحات المفضلة فى حين أن اللغة الطبيعية تتيح لنا حرية الاختيار) ويكمن السبب فى مرونة اللغة الطبيعية؛ فكما يمكن أن يكون لدينا أكثر من كلمة واحدة تدل على المفهوم نفسه، فكذلك يمكن أن يكون لدينا أكثر من تركيب نحوى يعبر عن الموضوع نفسه. مثال ذلك: يمكننا أن نقول: صناعة الأكياس الورقية، أو عملية الاتصال: "Paper sacks manufacture"

أو "the Communication process" "manufacture of paper Sacks" أو "cess of Communication، ومن الواضح أننا يمكن أن يكون لدينا شكل من أشكال علاقة التقابل ينشأ لا من الارتباط الدلالي بين الألفاظ ولكن من خلال ارتباطها التركيبي. وكما أننا نجد أن من المفيد أن نتحكم في المتقابلات الدلالية، فكذلك سوف نحقق الاطراد فقط إذا نحن تحكمتنا في المتقابلات التركيبية بالطريقة نفسها. وإن ما يعنيه هذا هو أننا يجب أن نقرر ما الترتيب الذي سوف نجمع المصطلحات عليه في خيط معين، وهكذا نحافظ على المعنى الصحيح ولكننا نستبعد الامكانيات البديلة.

إن نظام التوفيق Comhination order سمة كبرى في كل النظم المسبقة، رغم أنها لا ترد تحت هذا الاسم دائما. وهي في خطط التصنيف تعرف عادة باسم نظام ترتيب الأوجه citation order، في حين استخدم كوتس coates ترتيب الأهمية Significance order، لكي يؤكد على المبدأ الذي بنى عليه خيط مصطلحاته. وأيا ما كان الاسم، فإن المبدأ يظل هو هو: ضمان أن الموضوع المركب نفسه يعامل دائما بالطريقة نفسها، بصرف النظر عن الطريقة التي تعبر عنه بها اللغة الطبيعية.

إن عنصر الاطراد هذا هو في الغاية من الأهمية، ليس فقط من وجهة نظر المكشف الذي يبحث عن الإرشاد إلى الطريقة التي يعالج بها موضوعا معينا، ولكن أيضا من وجهة نظر المستفيد. بالنسبة للمكشف، تتضح أهمية ألا يكشف وثيقة بطريقة معينة إذا كان زميلة كشف وثيقة مشابهة بطريقة أخرى؛ وقد أثبت البحث أن الاطراد بين المكشفين بل وحتى بالنسبة للمكشف نفسه صعب جدا تحقيقه^(٤)، ولكن من المؤكد أن المشكلة تخف حدها إذا كانت هناك قواعد واضحة متبعة.

وبالنسبة للمستفيدين، فإن التأثير ذو وجهين. في المحل الأول، فإنه يعني أنهم إذا ما عرفوا كيف كُشِفَ موضوع ما، فإنهم سوف يجدون كل الوثائق المشابهة، التي من المفترض أن تحمل درجة ما من الصلة، إذا كانت كشفت

بالطريقة نفسها. وثانيا، فإنه يعنى القدرة على التنبؤ predictability : القدرة على أن نعرف سلفا كيف ستكشّف وتأتق تتضمن مفاهيم مختلفة، ولكنها فى نموذج أو إطار مشابه. وإن بعض الأمثلة سوف توضح هذا.

إحدى القواعد التى استخدمها كوتس هى أننا حينما يكون عندنا شيء thing عرّف بالمادة material التى صنع منها، فإن الشيء يسبق المادة. والمستفيد الذى يجد الرأس "Tanks, Aluminium"، ويعرف معناه، سوف يعرف فى مناسبة أخرى أن الرأس المطلوب هو "Bridges, concrete" هو. ولاتوجد مثل هذه القواعد فى قرمك؛ ما الذى يجب أن يفعله الجغرافى بالتسلسل تحت الكلمة Geographic ...? (المصطلحات المفضلة بالبنط الثقيل).

Geographic information systems

Geographic models

USE Geography, economic—Mathematical models

Geographical distribution of animals and plants

USE Biogeography

Geographical distribution of man

USE Human geography

Geographical distribution of plants

USE Phytogeography

Geographical location codes

Geographical models

USE Geography—Mathematical models

Geographical museums

Geographical myths

Geographical names

USE Names, geographical

Geographical photography

USE Photography in geography

Geography—models

USE Relief models

أين يجب علينا أن نبحث عن معلومات إستعمال النماذج models فى الجغرافيا؟ أين نجد المعلومات عن السمات الجغرافية؟ مثل هذه الأمثلة مرة

مدمرا فإذا كانت هناك إحالة تبادلية إلى الرأس المفضل، فإن هذا يقلل ولكن لا يلغى الاحباط عند عدم القدرة على أن نخمن تخميننا صحيحا لأول وهلة، وفي كل الحالات التي لا يكون فيها إحالة تبادلية، وما أكثرها، ماذا يمكن أن نفعل؟ إن الاطراد سمة مفتاح في لغة التكشيف الصحيحة.

بعض نظم التوفيق الموصى بها

كان كتر أول كاتب يناقش الترتيب الذى يجب أن تظهر عليه المصطلحات فى رأس لموضوع مركب، وقد أوصى باستخدام نظام اللغة الطبيعية «إلا إذا قررنا أن ثمة كلمة عدا الأولى أكثر أهمية منها؛ وبالنسبة لكتر تعنى كلمة مهمة "significant" معنى واضحا جدا فى هذا السياق، ولكن هذا المعنى ضاع بمرور السنين، وترتب عليه نتائج ناقشناها فى الفصل ٨.

وقد قدم كايزر Kaizer (قيصر) (الفصل ٨) الصيغة البسيطة المحسوس - العملية - process - Conerete، التى هى صحيحة أساسا، ولا زالت تستخدم بواسطة معظم النظم الحديثة، ولكنها لم تذهب إلى أبعد من ذلك. أما رانجاناثان فقد قدم خمس فئات أساسية (جوهرية) واعتبرها مسلمات فى نظامه: الشخصية Personality؛ المادة Matter؛ النشاط Energy؛ المكان Space؛ الزمان Time (نوقشت الفئات الخمس الأساسية فى الفصل الأخير)، ووضع صيغة ش م ط ن ز PMEST للترتيب المقتن، أو ش م ط ن ز للرووس الهجائية، وهى أشهر صيغة للتوفيق (أو ربما الأكثر سمعة سيئة، حيث تعتمد على وجهة نظر واحدة). وقد طور كوتس (الفصل ٨) نظام كايزر بحيث يعطى: الشئ - الجزء - المادة - الفعل - الأداة - Thing - Part - Material - Action - Agent، فى حين اقترح فيكرى (الفصل ٩):

Substance (Product) - Organ - Constituent - Structure - Shape - Property - Patient (Raw material) - Action - Operation - Process - Agent - Space -

Time المادة (المنتج) - العضو - المكون - البنية - الشكل - الخاصة - المتلقى
(المادة الخام) - الفعل - العملية - المعالجة - الأداة - المكان - الزمان .

ولحسن الحظ، فإنه لا ينبغي علينا أن نتذكر كل هذه الاجابات المختلفة -
والقائمة السابقة ليست كاملة - على المشكلة نفسها. وعند تحليل مجال
موضوعي معين، فمن الممكن أن نجد غالبا نظام توفيق في بنية الموضوع
نفسه، ولكن المقترحات المعممة تفيد في أنها تذكرنا بنوع الفئات التي علينا أن
نبحث عنها، والطرق التي يمكن بواسطتها توفيقها بصورة مفيدة. ويجب أن
نشير كذلك إلى أن نظاما مثبتا للتوفيق يمكن أن يثير مشكلات، وبصفة خاصة
في الترتيب المقنن، حيث التجميعات الناتجة من نظام التوفيق قد لاتناسب كل
مستفيدينا. وقد ناقشنا هذه المشكلات بتفصيل أوفى في الفصل ٩. وفي
التكشيف اللاحق، حيث توفيق المصطلحات لحظة البحث، لا توجد مشكلات
لنظام التوفيق، ولكننا من ناحية أخرى نفقد عنصر المعنى الذي يقدمه لنا تركيب
الخيط في المصطلحات المسيقة .

مشكلات الربط المسبق

ما المشكلات التي أشرنا إليها في الفقرات السابقة، وما المعايير التي يمكن
صياغتها لحلها؟ كما فعلنا بالنسبة للمشكلات التي ناقشناها حتى الآن، يمكننا أن
نقيس الحلول الجزئية (الخاصة) التي نجدها في لغات التكشيف على تلك
المعايير .

هناك شكلتان أساسيتان: الأولى هي أن اختيارنا لنظام التوفيق قد لايناسب
كل قرائنا، الذين قد يجدون أننا قد شتتنا على الرفوف أوفى الفهرس المفاهيم
التي هي الاهتمام الأول لهم. والثانية هي، كما رأينا في الفصل ٢، أي عرض
أو تمثيل خطى طولى للمعرفة متعددة الأبعاد يمكن فقط أن يعطينا وصولا إلى
مصطلح واحد، هو ذلك المصطلح الذي يأتي أولا في نظامنا للتوفيق ومن ثم
يكون عنصر الترتيب (التصنيف) Filing. وفي الحقيقة فإن هذين مظهران

مختلفان للمشكلة نفسها: حقيقة أن أى مصطلح بعد الأول يكون مخبأ، ولا يمكن أن نجده بصورة مباشرة. ويجب أن نلاحظ أن مشاكلنا هى هى سواء أعطتنا لغة التكشيف المختارة ترتيبا هجائيا أم مقننا؛ ومع ذلك فربما كانت أكثر حدة مع الترتيب المقنن، حيث أن الهدف الأكبر له هو إبراز العلاقات عن طريق تجميع الموضوعات المرتبطة معا بطريقة مفيدة. فإذا لم تبرز تجميعاتنا العلاقات التى يهتم بها مستفيدونا، فإنها ستكون عائقا أكثر منها مُعيناً. ولسوء الحظ، فإن علينا أن نقبل حقيقة أننا نادرا ما نجد نظاما للتوفيق يعطينا تجميعا للموضوعات المترابطة يرضى كل قرائنا، ويجب علينا فى العادة أن نسعد معظم الناس معظم الوقت.

وهناك ثلاث معالجات لهذه المشكلات. الأولى هى أن تعد مدخلا رئيسيا main entry واحدا (أى: رأس زائد تفاصيل عن الوثيقة) ومن ثم نوفر الوصول إلى هذه من نقط البحث الأخرى عن طريق الاحالات التبادلية. والثانية هى أن نعد مداخل موحدة unit: لكى نرتب نسخا من المدخل الرئيسى تحت أكبر عدد من نقط الوصول (الرؤوس) نظن أنها ضرورية. الطريقة الأولى اقتصادية، من حيث أن إحالة تبادلية واحدة فقط من الرأس غير المفضل إلى المفضل سوف تكفى مهما كان عدد المداخل التى يتضمنها الرأس المفضل. ومع ذلك فإن الطريقة الثانية أفضل من وجهة نظر القراء، من حيث أنها تعنى أنهم لن يحالوا من الرأس الذى فكروا فيه أولا إلى مجال فى مكان آخر.

ومن المهم أن نميز بين المدخل والاحالة التبادلية، حيث أن الفرق فى النظم اليدوية أو المطبوعة يمكن أن يكون مهما من الناحية الاقتصادية. ولنفرض أن لدينا وثيقة وواجهتنا قضية اختيار إعداد مدخل رئيسى واحد وأربع إحالات تبادلية، أو أربعة مداخل موحدة؛ إن كلا من الطريقتين تعطينا العدد نفسه من المداخل التى ذكرنا، رغم أن المداخل الموحدة الخمسة سوف تشغل حيزا أكبر من كشاف مطبوع. فإذا أضفنا إلى الفهرس تسع وثائق أخرى من الموضوع نفسه، فإن الطريقة الأولى سوف تعطينا عشرة مداخل رئيسية والإحالات التبادلية

الأربع نفسها، فى حين تعطينا الطريق الثانية ٥٠ مدخلا موحدا. فإذا كان عندنا ٢٠ موضوعا مشابها، فسوف تعطينا الطريقة الأولى ٢٠٠ مدخلا رئيسيا و ٨٠ إحالة، ولكن الثانية تعطينا ١٠٠٠ مدخلا موحدا. والفارق هو أن المدخل الرئيسى يحتوى على معلومات عن الوثيقة، ومن ثم يمكن أن ينسب أو يربط فقط مع هذه الوثيقة المخصصة، فى حين أن الإحالة تربط نقاط الوصول المترابطة ومن ثم تكون جيدة بالنسبة لأى شىء يتصل بنقاط الوصول هذه. ومع ذلك فإن كلا منهما نقاط مدخل فى التسلسل، ويجب علينا أن نوفر نقطة وصول تحت كل مصطلح ينتظر أن مستخدمه الاستفادة فى البحث، إلى المدى الذى يمكن أن نؤسس هذه النقاط سلفا.

وقد تغيرت اقتصاديات هذا الوضع على مر السنين. ففى أيام الفهرس المطبوع (وفى الأيام الأولى للفهارس المطبوعة المنتجة بواسطة الحاسب) كانت تكلفة المداخل الإضافية عالية، ولذلك استخدمت الاحالات التبادلية، أو ربما استخدمت المداخل المختصرة. ولما كانت الفهارس البطاقية قد أصبحت النموذج، فإن المداخل الموحدة قد أصبحت أكثر شيوعا، وبصفة خاصة مع دخول البطاقات المطبوعة كتلك التى أصدرتها مكتبة الكونجرس؛ ومع ذلك بقيت التكلفة العالية لتصفيف وتخزين البطاقات.

والمعالجة الثالثة جاءت مع دخول تسجيلات مارك والفهارس المبنية على الحاسب. مع قاعدة البيانات على الحاسب، يوجد تسجيلة واحدة فقط مع نقاط إتاحة متعددة، وتوقف من ثم تمييز الطرق عن بعضها. ويقترح أو برين 'O Brien^(٥) أن «الإتاحة عن طريق الكلمة المفتاح هى مُحَرَّرٌ ومنتج جانبي للإسترجاع المحسب، فى حين أن النظم المقيدة هى من تراث الفهارس اليدوية التى هى نظم استطلاعية ذات بعدين». ومع ذلك فلا زالت توجد بيليوجرافيات مطبوعة كثيرة، تكلفة إنتاجها عالية، ولازلنا نجد أن استخدام الإحالة التبادلية يقود إلى مدخل رئيسى مفرد.

ويجب أن نفرق أيضا بين المصطلحات التي توجد في لغة الكشف وتلك التي توجد في لغة المداخل. ومن المعتاد أن نعد إحالة استخدام (استعمل) USE من الأخيرة إلى المصطلحات المفضلة؛ أما المصطلحات غير المفضلة فليست نقاط وصول. وهنا أيضا، يمكن لنظام محاسب متطور أن يأخذنا أو توماتيا من مصطلح غير مفضل إلى نقطة الوصول المفضلة.

الاحالات التبادلية

في جدول ٦,٤ ذكرنا الموضوع: معالجة الألومنيوم بالحرارة؛ وهذا الموضوع يدخل في معظم لغات التكشيف كما يأتي:

Aluminium - Heat treatment

وهنا استخدمنا المبدأ المتفق عليه وهو أن الشيء (المحسوس) أهم من الفعل (العملية)، ومن ثم فإنه يجب أن يسبقها في نظام التوفيق الذي يحدد شكل خيط المصطلحات الذي نبينه. وبالنسبة لهذا الرأس خاصة، فإن هناك طريقة لبناء الاحالات تفرض نفسها وهي.

Heat treatment - Aluminium USE Aluminium - Heat Treatment

وعن طريق قلب المصطلحات، يمكن أن نعد إحالة تقابل. ويمكن أن نعد بصورة تبادلية إحالة رتبية:

Heat treatment NT Aluminium - Heat treatment

ولكن كيف يمكن أن نستخدم هذا المدخل المباشر إذا كنا نكشف مقالة عن:

"Ethnic minority adolescents identity in

the new Europe: atrans - cultural approach [using psychotherapy]"

إن الرأس الذي نجده لهذا الموضوع من ASSIA هو:

Ethnic groups - Young people - Ethnic identity - Psychotherapy - Cultural aspects.

مثل هذا الرأس يحاول أن يمثل الموضوع بصورة كاملة، وهو يعطى مدخلا مفردا مخصصا، ولكنه يثير بشكل حاد المشكلات التي نتناولها. فى المحل الأول، رغم أننا نأمل أن يمكن الترتيب حسب الأهمية المستفيدين أن يجدوا هذا بصورة مباشرة، فليس هناك ضمان بأن هذا سوف يكون ناجحا فى كل مرة، ويجب أن نعمل حسابا للمستفيد الذى يبحث تحت واحد من المصطلحات المخبأة. وفى المحل الثانى، فإن المستفيدين الذين يهتمون بالموضوع "young people" قد يرغبون فى أن يعرفوا عن هذه المقالة التى تعالج مراهقى الأقليات العريقة.

طريقة السلسلة

وضع انجانا ثان طريقة، هى طريقة السلسلة Chain Procedure، يمكن بواسطتها أن نعد العدد الأدنى من الإحالات، ومع ذلك نظل متأكدين أن القراء سوف يجدون نقطة مدخل تحت كل مصطلح مهم فى مثل هذا الموضوع المركب. وهذه هى فى الحقيقة الطريقة التى استخدمها ديوى لبناء «الكشاف النسبى» لتصنيفه، ولكن رانجانا ثان وضع لأول مرة أساسها النظرى. فى هذه الطريقة نبدأ بكتابة «سلسلة» المصطلحات:

Ethnic groups - Young people - Ethnic identity - Psychotherapy - Cultural aspects

ولكى نبني الإحالات التبادلية، نكتب المصطلح الأخير فى السلسلة، يليه كل مصطلح من المصطلحات السابقة عليه بدوره؛ وهذا يشكل المدخل تحت ذلك المصطلح ويقودنا إلى الرأس المفضل بشكل كامل:

Cultural aspects - Psychotherapy - Ethnic identity - Young people - Ethnic groups

See

Ethnic groups - Young people - Ethnic identity - Psychotherapy - Cultural Aspects

ويمكننا الآن أن نتقل إلى المصطلح قبل الأخير، ونسير على الطريقة نفسها:

Psychotherapy - Ethnic identity - Young People - Ethnic groups

See

Ethnic groups - Young people - Ethnic identity - Psychotherapy

وعن طريق مداخل مشابهة لكل مصطلح مهم يمكننا أن نضمن أن يكون بإمكاننا الوصول إلى جزء من الرأس على الأقل، بدءاً بالمصطلح الرائد، مهما كان المصطلح الذي نفكر فيه. فإذا كان هناك n من المصطلحات المهمة، إذن فإننا نحتاج $(n-1)$ نقط وصول إضافية.

وإن إنتاج ASSIA بالحاسب يعنى أن كل مدخل فيه يمكن أن يرقم. وترتب المداخل بواسطة الرأس الكامل، فى حين يبنى الكشاف الموضوعى بواسطة طريقة السلسلة لكى يقود إلى رقم (أرقام) المستخلص المناسب وليس الرأس. ونجد مداخل الكشاف التالية بالنسبة لهذا المثال :

Cultural aspects - Psychotherapy - Ethnic identity - Young People - Ethnic groups 4333

Psychotherapy - Ethnic identity - Young people - Ethnic groups 4333

Ethnic identity - Young people - Ethnic groups 4333

Young people - Ethnic groups 4333

Ethnic groups 4325 - 4333

ونجد كذلك إحالة تبادلية للمصطلح المرتبط :

Adolescents

See Also

Yong people

وهذا يعطينا نظام «مدخل رئيسى» يتألف من الرأس الكامل، والمرجع الببليوجرافى والمستخلص، يسندها كشاف موضوعى يمكن الوصول إلى كل

مصطلح مهم فى الرأس، وكشاف مؤلف. وإن استخدام أرقام المستخلصات وليس الرؤوس يتجنب واحدة من مشكلات طريقة السلسلة : قد تقودنا إلى رؤوس غير موجودة، أو إلى رؤوس أوسع مما نريد. مثال ذلك: يقودنا الرأس : الجماعات العرقية إلى مستخلصات متعددة، لأنه رأس أوسع من الرأس المستخدم لهذا المستخلص بالذات. ولن نجد الرأس :

Young people - Ethnic groups

إلا فى الترتيب المعكوس : Ethnic groups - Young people باعتباره جزءا من الرأس الأكثر تخصيصا لهذا المستخلص. ويشبه ذلك، المستفيد الذى يبحث فى كشاف تعد تحت الهندسة فى حين أن الذى يريده هو هندسة الاتصالات هذا سوف يجعله بحاجة إلى أن يبحث فى الجداول والملخصات لكى يحصل على مايريد خلال الأرقام من ٦٢٠ إلى ٦٢٩٠٤.

إن طريقة السلسلة بسيطة ويمكن تحسينها بسهولة. ومن المنطقي كذلك أنها يجب أن تعطى الحد الأدنى من أعداد المداخل الاضافية التى تقود إلى الرأس المطلوب. ومع ذلك، ففى الأوضاع التجريبية، مثل مشروع EPSILON^(٦). ظهر أن طريقة السلسلة لم تعط الاقتصاد المتوقع فى المداخل. والسبب فى هذا ليس واضحا، ولكنها نقطة يجب أن نضعها فى أذهاننا. وربما تولدت إلى حد ما من النتائج التى توصلت إليها بوب BNB؛ إن استخدام تكشيف بريسييس بدلا من كشاف السلسلة الذى استخدم من ١٩٥٠ إلى ١٩٧١ كان ينتظر أن يقود إلى زيادة فى حجم الكشاف، ولكن هذا لم يحدث. وقد يكون من المهم أن نبحث الأسباب التى تجعل تكشيف السلسلة لايرقى إلى مستوى التوقعات.

الدوير والتبديل

رأينا فى الفصل ٣، أن من الممكن أن نبنى كشاف كويك KWIC عن طريق الإبقاء على الموضع الترتيبى للمصطلح المفتاح ثابتا، ولكن مع تحريك الكشاف كله عبر الصفحة بحيث تظهر كل كلمة فى الموضع الترتيبى بدورها.

مثال ذلك: يمكننا أن نأخذ رأس ASSIA الميين فيما سبق ونبنى كشافا من نوع كويك:

| | | |
|-----------------------------------|---|------|
| Ethnic identity - Psychotherapy - | Cultural aspects Ethnic groups - Young people | 4333 |
| Psychotherapy-Cultural aspects | Ethnic groups-Young people-Ethnic identity | 4333 |
| Ethnic groups-Young people- | Ethnic identity-Psychotherapy-Cultural aspects/ | 4333 |
| Young people-Ethnic identity- | Psychotherapy-Cultural aspects Ethnic groups | 4333 |
| Cultural aspects/ Ethnic groups | Young people-Ethnic identity-Psychotherapy | 4333 |

هذا النوع من الكشاف هو كشاف مدور rotated index^(٧)، وكما سبق أن ذكرنا في الفصل ٣، فإنه يوجد أكثر ما يوجد كمكمل للمكانز، لكي يكشف عن الكلمات المخبأة في الرؤوس التي تتألف من عدة كلمات.

والتبديل permutation* مصطلح رياضي يحيل إلى بناء كل الترتيبات الممكنة لمجموعة معينة من العناصر؛ إذا كان عندنا ثلاثة عناصر هي A,B,C إذن فإن هناك ستة ترتيبات أو تباديل ممكنة:

ABC; ACB; ACA; BAC; CAB; CBA

إن عدد التباديل لـ n من الأشياء هو n!

(n factorial = n x n - 1 x n - 2 ... 3 x 2 x 1)

وعلى هذا فإننا بالنسبة للرأس الذي نبخته أعلاه يمكن أن يكون عندنا 5! أو ١٢٠ مدخلا ممكنا، علاوة على الخمسة التي بنيناها باستخدام التدوير. ولأسباب غير معروفة، فإن الكشافات المدورة rotated في المكانز يطلق عليها غالبا الكشافات المبدلة permuted، ولكن هذا الاسم غير صحيح ويجب تجنبه.

* التباديل permutations والتوافيق combinations هما مصطلحان رياضيان وقد استخدمهما رانجاناثان الذي بدأ حياته أستاذا للرياضيات قبل أن يدخل إلى عالم المكتبات في سنة ١٩٢٤ (المترجم).

المدخل المتعدد Multiple entry

تناولنا حتى الآن الوضع الذي نعدفيه مدخلا رئيسيا واحدا يدعمه كشافات من أنواع مختلفة. والبديل هو أن نعد مداخل موحدة، حيث يرتب المدخل الذي يتضمن تفاصيل كاملة عند كل رأس مناسب. وقد أخذنا مثالنا على طريقة السلسلة والذي أعطيناها. سابقا من المطبوع البريطاني ASSIA؛ والمدخل الموحد أكثر شيوعا في الكشافات الأمريكية مثل PAIS International. مثال ذلك، يوجد المدخل التالي تحت ثلاث رؤوس :

Seeger, N. Liechtenstein : the differential tax treatment of offshore and domestic enterprises. Bul Internal Fiscal Docum, 48 579-81 N 1994 [Four - line abstract].

والرؤوس والرؤوس الفرعية هي

International business enterprises - location

Liechtenstien - Tax policy

Taxation - Liechtenstein

وهنا أيضا إحالة تبادلية :

Business enterprises

See also

International business enterprises

ويمكن أن نعقد مقارنة مشابهة بين :

(بريطاني) Library and information science abstracts وبين

(أمريكي) Library Literature. وكل منهما يغطي المقالة الآتية :

Klein, S. S. "Your right to privacy and the AIDS virus : : a selective bibliography", Legal reference saervices quarterly, 13 (3) 1994, 115-127.

في ليزا LISA هناك مدخل واحد 94-2955، مع مستخلص، تحت الرأس :

تنمية المجموعات Collection development. وفي الكشاف الموضوعي نجد :

Privacy – Patients – Acquired Immune Deficiency Syndrome – Bibliographies
– Selection aids – Acquisitions 2955
Patients – Acquired Immune Deficiency Syndrome – Bibliographies –
Selection aids – Acquisitions 2955
Acquired Immune Deficiency Syndrome – Bibliographies – Selection aids –
Acquisitions 2955
Bibliographies – Selection aids – Acquisitions . . . 2955 . . .
Selection aids – Acquisitions . . . 2955 . . .
Acquisitions . . . 2955 . . . [twelve blocks of entries, one for each month]

ومدخلا لـ Klein, S.S. فى الكشاف المستقل للمؤلف .
فى Library Literature نجد ثلاثة مداخل وإحالة تبادلية فى التسلسل
المعجمى المفرد (الواحد) :

Right of privacy – Bibliography
Legal literature – Bibliography
Klein, Sandra S.
Bibliography

See also subhead Bibliography under specific subjects and under
specific types of literature.

العلاقات الجانبية

افتراضنا حتى الآن أن من الممكن بناء خيط من المصطلحات، باستخدام
العلاقات التركيبية أو بصورة صريحة فى شكل دالات الدور، وذلك لإنتاج تمثيل
مطرود لأى موضوع بسيط أو مركب (أى موضوع يشتمل على مفهوم واحد فقط
أو موضوع يشتمل على مفاهيم متعددة)، ثم نتداول هذا الخيط تبعا لطرق
مباشرة مختلفة لكى نمكن المستفيدين من أن يجدوها ليس من خلال المصطلح
الذى يعبر عن المفهوم الذى نعتبره الأهم فقط، ولكن أيضا من خلال أى من
المفاهيم الأخرى. قد لاندل بصورة صريحة على الدور أو الفئة التى يقع فيها
مصطلح معين، ولكننا نفترض أن هذا سوف يكون واضحا من السياق. وإن
الأمثلة التى أوردناها حتى الآن تبرر فى الحقيقة هذه الافتراضات.

فإذا تناولنا جدول ٧, ١، فإننا يمكن أن نرى أننا لم نوفر بعد لأنواع العلاقة المعروضة. والمصطلح الذي يستخدم للدلالة على هذه العلاقة، هو العلاقات الجانبية Phase relationships، وقد صكه رانجاناثان، ويمكننا أن نميز أربعة أنواع خاصة^(٨). في الجماعة الأولى نجد موضوعات تعالج وفي الذهن جمهور معين؛ والموضوع نفسه يظل هو هو، ولكن الأمثلة التي تستخدم لتوضيحه سوف تجيء من الموضوع الذي يهتم به الجمهور المقصود. قد يكون كتاب تقديمي عن الإحصاء كتب للعاملين في المكتبات - قد يكون ذا فائدة بالنسبة لطلاب إدارة الأعمال، ولكنه لا يتتظر أن يكون ذا فائدة كبيرة لطالب يريد كتابا عن خدمة المراجع مثلا. هذا النوع من العلاقة الجانبية يطلق عليه جانب الميل bias phase، والموضوع الذي يعالج هنا هو موضوع الكتاب، وليس الجمهور المقصود.

في الجماعة الثانية، نجد موضوعا يتأثر بآخر، ويسمى جانب التأثير in-fluence phase، ويكون الموضوع المتأثر هو المقصد (المحور) Core. وفي الثالثة، جانب العرض أو الشرح expository phase، يكون الموضوع المشروح، وليس الموضوع «الأداة» هو الموضوع الأهم. وفي الجماعة الرابعة والتي تبين المقارنة comparison أو التفاعل interaction، نجد أن الموضوعين معالهما أهمية متساوية.

جدول ٧, ١ العلاقات الجانبية

-
- 1 Middle management for nurses
 - 2 Fluid mechanics for civil engineers
 - 3 Basic statistics for librarians
 - 4 Methylmercury in fish: effects on human health
 - 5 The literary impact of *The golden bough*
 - 6 The effect of PAS on the resistance to streptomycin of tubercle bacilli

- 7 Typewriter behaviour: psychology applied to teaching and learning typewriting
 - 8 Literature through art: a new approach to French literature
 - 9 Hamlet: an analytic and psychologic study
 - 10 Science and politics
 - 11 Church and state
 - 12 Religion and science
-

وحيثما نحاول أن نكتب الخيوط التي تمثل العلاقات الجانبية بصورة كاملة، نجد أننا مضطرون إلى أن نجعلها صريحة باستخدام مصطلحات تدل على نوع العلاقة المتضمنة. مثال ذلك، الرأس:

Aluminium - heat treatment

لايستخدم حروف الجر، ومع ذلك فإن قليلا من القراء فقط هم الذين سيسيثون فهمه. وحتى الموضوع الأكثر تعقيدا والذي اقتبسناه سابقا:

Ethnic groups - Young people - Ethnic identity - Psychotherapy - Cultural aspects

لايحتاج إلا إلى جهد قليل لفهمه. وعلى العكس، إذا حذفنا الكلمات المتفرعة، فإن الرؤوس:

Middle management - nurses

English literature - *The golden bough* (instead of *The golden bough* - influence on English literature)

French literature : art (instead of French literature - exposition through art)

Science : politics

إما أن تكون متعددة المعنى أو مضللة. ويمكن تعرف العلاقات الجانبية بواسطة هذه الخاصية: الحاجة إلى جعل العلاقة صريحة أو مباشرة - في النظام الهجائي بواسطة استخدام الكلمات التي تعبر أو تدل على أفكار مثل التأثير، المفعول أو النتيجة، الجمهور، المقصود، التفاعل، المقارنة، وفي النظام المصنف باستخدام الرمز مثل ٤٢ - ... في تعدد. وهناك نقطة أخرى عن العلاقات الجانبية من الضروري فهمها. ذلك أنها بالضرورة علاقات خاصة؛

ورغم أننا نعرف سلفاً أن كتباً سوف تكتب يعالج فيها موضوع واحد مقارنة بآخر، أو أن موضوعاً معيناً تأثر بآخر، أو تأثر بجمهور معين في الذهن، رغم ذلك فإننا لا يمكن أن نعرف أى الموضوعات سوف تعالج بهذه الطريقة. فالعلاقات الجانبية لا يمكن التنبؤ بها بالطريقة نفسها التي يمكن بها أن نتوقع، مثلاً، أن المعادن سوف تخضع للأشكال المختلفة من المعالجة، أو أن المحاصيل سوف تتأثر بالآفات. وكل ما يمكننا عمله هو أن نوفر بطريقة كافية لهذه العلاقات عن طريق تعريف إمكان ورودها.

في الأنواع الثلاثة الأولى المبينة في جدول ٧١، يمكن أن نتعرف واحداً من المفهومين باعتباره المفهوم المفتاح. وفي جانب الميل، حيث يعالج الموضوع من زاوية جمهور معين، يكون الموضوع المعالج هو الجانب الأول؛ وفي جانب التأثير، يكون الموضوع المتأثر، وفي وجه الشرح، يكون الموضوع المشروح. في هذه الحالات الثلاثة يمكن أن نبني خيطاً من المصطلحات التي تمثل الموضوع بشكل غير متعدد المعنى، ويمكن أن نبني الإحالات التبادلية عن طريق تغيير الكلمات الرابطة أو الترتيب. وهكذا يمكن أن يكون لدينا:

Statistics for librarians

Librarians - statistics for

The golden bough - influence on English literature

English literature - influence of the golden bough

French literature - expounded through Art

Art - expounding French literature

ومع ذلك، فإن الطريقة الوحيدة التي يمكن بها أن نعالج علاقة التفاعل، التي يمكن أن ينظر إليها على أنها تأثير ذو طريقتين، هي أن نعالج المفهومين المتضمنين على أنهما متساويان في الأهمية.

Science in relation to Politics

Politics in relation to Science

ويمكن أن نستخدم الكلمة *and*، كما في *Science and Politics*، ولكن هذا قد يكون مضللاً. فإذا كان لدينا كتاب دراسي نموذجي عن الفيزياء عنوانه

Mechanics, heat, and sound، فإننا يمكن أن نكون واثقين أن هذا ربط مادي لثلاثة أعمال منفصلة داخل غلاف واحد، كإجراء اقتصادي، وليس أنها تعرض أى نوع من التفاعل. وإذا كان لدينا عمل عن تعدين الفضة والرصاص، فإن عندنا فى الحقيقة موضوعين منفصلين : تعدين الفضة، وتعدين الرصاص. وحينما كتب كوينتين هوج Science and politics : Quintin Hogg، فإنه كان يكتب عن تفاعلها : الأثر الذي قد تحدثه القرارات السياسية على العلم، مثال ذلك عن طريق تحديد أى المشروعات تتلقى التمويل، والعكس، الأثر الذي يمكن أن يتركه العلم على السياسة، مثال ذلك، عن طريق تمكين الساسة من أن يصلوا إلى جمهور باتساع العالم كله من خلال التلفزيون. ومع ذلك، فقد يكون عندنا عمل يعالج «علم السياسة» يناقش الطرق التي تصنع بها القرارات وتنفذ. وقد تكون هذه علاقة دائمة، وليست نوع علاقة التفاعل التي كنا نناقشها.

وفى ختام هذه المناقشة عن العلاقة التركيبية، فقد يكون جديرا بنا أن نقدم كلمة تحذير. فقد أشار جاردن، إلى أن التراكيب والدلالات تمضيان معا، وينبغي أن ينظر إليهما على أنهما مظهران لعملية صياغة عبارات واضحة وغير متعددة المعنى للموضوع وعلاقاته. وقد تعبر جملة تعبيريا كاملا عن كل مقتضيات التركيب ومع ذلك تكون لامعنى لها. وقل نقل أوستن^(٩). مثال شومسكى Chomsky المشهور : "Colourless green ideas sleep furiously". فإذا أخذنا الكلمات نفسها وغيرنا تركيبها، فإننا سوف تعطى معانى مختلفة كلية : ولننقل واحدا من أمثلة أوستن نفسه، "pollution of oil by water" فمن المؤكد أن معناها يختلف عن "pollution of water by oil"، رغم أن الكلمات المستخدمة متماثلة.

الغرض من العلاقات التركيبية

رأينا أن الغرض من العلاقات الدلالية هو تمكيننا من تحسين أداء الاستدعاء فى نظامنا. وفى المقابل، فإن الغرض من العلاقات التركيبية هو تمكيننا من

تحسين أداء الصلة، عن طريق السماح لنا بأن نخصص الموضوعات بصورة أدق. ويمكننا أن نفعل ذلك عبر العلاقات الدلالية إلى حد ما، عن طريق إدراك أن هناك مصطلحا أكثر تخصيصا من المصطلح الذي فكرنا فيه أولا، ولكن أقوى الطرق لزيادة الخصوصية هي من خلال الربط coordination. ويجب أن تقاس الخصوصية في أية لغة للتكشيف ليس من خلال مفرداتها، ولكن من خلال القدر الذي تسمح به من الربط كذلك. وقد رأينا في الفصل ٥ أن الربط عبر الاستخدام غير المحكوم لعلاقة AND البولية يمكن أن يقود إلى الفشل، ولكن في نظام للتكشيف المعين، تربط المصطلحات حينما تكون مناسبة ومناسبة فقط. ويمكن أن نستخدم مثلا سبق استخدامه : إذا وجدنا Youn peo - Ethnic groups بين مصطلحات التكشيف، فسوف نعرف أننا سوف نجد معلومات عن طريق توفيق هذه المصطلحات.

النظم المغلقة والنظم المفتوحة

إذا أردنا أن يكون استخدامنا للغة التشكيف مطردا، فيجب أن نسجل في مكان ما القرارات التي نتخذها، مثال ذلك عند اختيار المترادفات المفضلة أو نظام الأهمية. والسجل الذي يجمع على هذا النحو يسمى سجل الاستناد authority file، وقد يكون السجل من صنعنا نحن وقد يكون مبنيا بدرجة تزيد أو تنقص على قائمة مطبوعة. ولأغراض التوحيد. فمن المعتاد بالنسبة للمكتبات أن تستفيد من القوائم المطبوعة، وفي الحقيقة فإن الجزأين الثالث والرابع من هذا الكتاب قد خصصا لوصف مثل هذه الخطط. ومع ذلك، فقد تظهر مشكلة لخصها رانجاناثان في عبارة : الاستقلال الذاتي للمصنف.

فالمعرفة، كما أكدنا، ليست ثابتة، وتحتاج لغة التكشيف إلى المراجعة المستمرة إذا أرادت أن تظل حديثة. علاوة على ذلك، فإن الخطط العامة لما كانت ترتبط بالسند الأدبي لمجموعة معينة، أو لمجموعة افتراضية، فإنها قد تستبعد بعض الموضوعات التي نجدها متمثلة في مجموعتنا. ماذا نفعل إذا

وجدنا أن هذه هي الحالة؟ هل يمكننا أن نضيف بأنفسنا الرأس الجديد في سجل الاستناد، أم هل يجب أن ننتظر حتى ينتج الجامعون طبعة جديدة أو نشرة تغييرات؟ وإذا كانت الحالة هي الأولى، فإنه يقال عن النظام إنه مفتوح open، في حين أن النظام في الحالة الثانية يكون مغلقا closed. وقد أدرك ديوى المشكلة منذ مرحلة مبكرة، واقترح أنه إذا لم يوجد موضوع معين في خطته، فإننا يجب أن نصنف الوثيقة في أقرب رأس يشتمل عليها، وهو رأى مبنى على نظرية أن الموضوعات الجديدة تنشأ عادة عن تجزئ أو عن تفتيت، أو انشطار "fission" الموضوعات الموجودة فعلا. ثم يقوم الجامعون بتوفير مكان للموضوع الجديد في الطبعات التالية، وهو أمل متفائل لم يتحقق دائما عند التطبيق. وكما رأينا في الفصل ١، فإن نمو الموضوعات الجديدة عن طريق الانشطار كان معتادا في الحقبة (العصر) التي كان التوجه فيها نحو العلم أو المجال، ولذلك كان حل ديوى معقولا تماما في ذلك الوقت - رغم أن من العدل أن نشير أن هذا الحل لم يوفر للعلاقات الجانبية، وبخاصة تلك التي لم يكن من الممكن فيها تحديد الموضوع المفتاح. إن حل «أقرب الرؤوس المشتملة على الموضوع» يضع المكشف أمام اختيار مستحيل هو أن يحاول أن يجد مكانا لموضوع جديد نما عن طريق «اندماج fusion» علوم مستقلة سابقة، في حين أن المعالجة متعددة المجالات (العلوم) لعصر التوجه نحو الرسالة قد اتضح أنه قد افتتح من خلال معالجة مبنية على الفصل غير متعدد المعاني للعلوم المستقرة.

ولما كان الترتيب الهجائي واضحا بذاته، فإن إضافة الرأس الجديد في قائمة هجائية أسهل من إضافته في قائمة مرتبة ترتيبا مقننا؛ وكل ما نحتاجه هو معرفة شكل الرأس، مثل : هل نستخدم شكل المفرد أم الجمع من الاسم.

وتشتمل القوائم الهجائية مثل قرمك أو سيرز على قواعد لإضافة الرؤوس

الخاصة بالأفراد المعروفين لنوع ما، مثل الأشخاص، الزهور، الأماكن، الحيوانات. وهناك أيضا الرؤوس النموذجية وهي التي تبين النموذج الذي يجب أن نبني عليه الرؤوس التي تتناول مثلا أعمالا عن مؤلفين مخصصين وكتبهم.

وإنه لأصعب من ذلك أن نؤسس سلفا الرمز الصحيح لموضوع جديد في خطة التصنيف، حيث أننا هنا نتعامل مع لغة اصطناعية، ولم يقدم حتى الآن حل مرض. وتقدم وسائل التذكر الجذرية Seminal mnemonics التي ناقشناها في الفصل ٢١، حلا مؤقتا، ولكنه حل يكون الترتيب فيه مقرا بواسطة الرمز وليس من خلال حاجة الموضوع. ومن أجل هذا السبب فإن طريقة المراجعة التي تتبناها خطة التصنيف تكون مهمة، وقد ناقشنا هذه النقطة في كل خطة من الخطط التي درسناها في الجزء الثالث.

الحصر والتركيب

إن علينا في لغة التكشيف المسبقة أن نوفر لكل من المفاهيم المفردة وللموضوعات المركبة، ولهذا تصبح مشكلات النظام المغلق أكثر حدة. والأنظمة القديمة تشمل أو تحصر enumerate كل الموضوعات التي تبدو مناسبة، سواء في ذلك المفاهيم المفردة أو المركبة، ولاترك للمصنف الفرد أى استقلال ذاتي لإضافة الموضوعات الجديدة التي قد تظهر في مجموعة معينة.

وأما الخطط الأحدث فإنها تستخدم التركيب synthesis للتعبير عن الموضوعات وهذه الخطط تسجل المفاهيم المفردة فقط، ولكنها توفر للمصنف قواعد لبناء الرؤوس للموضوعات المركبة. ومن الواضح أن الطريقة أقوى بكثير من طريقة الخطط الحاصرة، ولكنها تعتمد على المفاهيم المفردة التي تسجلها الخطة، أو تكون ثمة قواعد لبناء الرؤوس منها. وهكذا فإن اللغة التي تعتمد على التركيب مفتوحة بدرجة أكبر بكثير جدا من اللغة الحاصرة، ولكنها قد لاتعطي المكشف الاستقلال الذاتي الكامل. ومن أجل هذا السبب فإن كثيرا من المكتبات المتخصصة التي تعالج الموضوعات الجديدة التي تنشأ من البحث قد

تجنبت أن تكون مقيدة بقائمة مطبوعة قد تكون قديمة وفضلت أن تبني لغات خاصة بها للتكشيف، سواء مقننة أم هجائية.

وكما أكدنا، فإن العلاقات الجانية هي أساسا خاصة : ولذلك فإن من الصعب جدا أن نعالجها في خطة حاصرة. ولا يمكن لخطة حاصرة أن تحصر إلا تلك العلاقات التي تكون ظهرت بالفعل، فنحن نجد في تعدد : العلم والدين، لأن هذا كان موضوعا لعدد من الكتب في الوقت الذي بنى فيه ديوى تصنيفه في ١٨٧٦، وقد ظهر في عالم النشر، من بين أعمال أخرى مهمة، كتاب دارون Darwin : أصل الأنواع origin of species في سنة ١٨٥٩ - ولكننا سوف نبحت عبثا عن «العلم والسياسة». ويمكن للخطة التركيبية أن توفر بسهولة أكثر للعلاقات الجانية.

ولا يوجد بطبيعة الحال خط فاصل حاسم بين النوعين من الأنظمة؛ قد نتوقع أن نجد مدى كاملا بين الخطة الحاصرة المغلقة حيث لا يوجد للمصنف حرية على الاطلاق، والخطة التركيبية المفتوحة التي تسمح للمصنف بأن يخصص أى موضوع حينما يظهر. ومعظم الخطط تقع في مكان ما بين طرفي النقيض، وربما كان من الضروري أن نذكر أنه بالرغم من أن إحدى المزايا المزعومة للخطط اللاحقة أنها أكثر حرية من الخطط المسبقة، فإن هذه ليست هي الحالة بالضرورة. فإن لغة كتلك التي تستخدم في ASSIA يمكن أن تكون لامحدودة، في حين أن بعض المكانز تقدم لغة محدودة بشكل صارم.

التصنيف النقدي

المكشوفون وسائط بين المؤلفين والمستفيدين : إلى أى حد يكون من حقهم فرض آرائهم أو أهوائهم الخاصة؟ والإجابة المباشرة هي : ليس من حقهم على الإطلاق، ولكن عند التطبيق يكون تحقيق هذا في الغاية من الصعوبة. إن لغة التكشيف تعكس لامحالة الخلفية الاجتماعية والثقافية لجامعها؛ وفي حين أنها تخدم المستفيدين الذين يتشابهون في الآراء فإن هذا ليس بالضرورة شيئا سيئا؛

ولكن حينما نحاول فرض ثقافة معينة على مستفيدين لهم ثقافة أخرى، هنا تنشأ المشكلات. والأمثلة على الشوفينية معروفة، رغم أنه من حسن الحظ أن هذه قد حذفت من قرمك وسيزر:

World War II : Germany's Conquest of Europe, Rescue by the United States (BCI)

أو نجد فروقا حقيقية في المعالجة:

قارن جداول الحكومة government في كل من تمك أو تك مثلا أو التربية education في تعع (وجهة النظر الأوربية وفي تمك) (وما) أو الفروق المثيرة للاستياء.

Disabled USE Handicapped (LCSH)
Disabled students USE Handicapped students
Disabled veterans USE Veterans, Disabled

أو الغطرسة العقلية

821.9 Minor poets [e.g. John Donne!] (early editions of DDC)

أو لطف التعبير

Falkland Islands

SN As a geographic subdivision this heading is used directly (LCSH)

وهذا يتجنب بصورة كاملة قضية ما إذا كان المدخل الموحد غير المباشر يجب أن يشتمل على الأرجنتين أو بريطانيا العظمى!

وأيا ما كان السبب، فإن التصنيف النقدي critical classification - فرض وجهة نظر المصنف على المستفيد - يوجد بأكثر مما نريد أن نرغب في الاعتراف^(١٠). وحتى خطط مثل تعع، التي لها سياسة دولية موحدة، ليست بريئة من الانحياز.

فى الطبعات التى سبقت الطبعة العاشرة، ١٩٨٦، كانت قرمك تخلو من رزس عن محمد (صلى الله عليه وسلم)؛ وعند المدخل : الاسلام كان هناك ملحوظة : تدخل هنا الأعمال عن الدين الذى محمد (صلى الله عليه وسلم) نبيه. فى الطبعة ١٠، أضافت قرمك الرأس : Muhammad d. 632، وهكذا تغضب جزءا كبيرا من العالم الإسلامى من خلال تجاهل وضعه كنبى. وفى الطبعة ١١ من قرمك غير الرأس إلى : Muhammad, Prophet, d. 632. ويجب أن يحذر المكشف من ألا يقدم الانحياز بدون علم أو دراية، وأن يدرك المستفيد أن الانحياز يمكن أن يوجد، ذلك إذا أردنا أن نحقق الهدف من جعل المعلومات متاحة لكل من يطلب.

لغات التكشيف

يطلق على النظام الذى يسمى الموضوعات بالطريقة التى وصفناها : لغة التكشيف indexing language، وهى، مثل أية لغة، تتألف من جزأين: المفردات vocabulary والتراكيب. فإذا استخدمنا المصطلحات كما تظهر فى الوثائق دون تعديل، فإننا نستعمل اللغة الطبيعية natural language وقد رأينا أن هذا يمكن أن يقود إلى مشكلات. وبخصوص المفردات، فإن المؤلفين المختلفين يمكن أن يستخدموا مصطلحات مختلفة للتعبير عن الفكرة نفسها - المترادفات - والتى تقود إلى تناقص فى الاستدعاء. أما بالنسبة للتراكيب، فإن الفكرة نفسها يمكن التعبير عنها بأكثر من طريقة واحدة :

child psychology أو psychology of children;

adult education أو education of adults

ولهذه الأسباب، فإن نظم التشكيف تمارس نوعا من التحكم على المصطلحات المستخدمة : أى أننا نستخدم لغة مقيدة Controlled vocabulary. كذلك فإننا نصوغ التراكيب المرنة للغة الطبيعية عن طريق السماح بإنشاءات أو صيغ معينة فقط : فبدلا من heat treatment of aluminium نستخدم :

libraries for children aluminium - heat treatment وبدلا من أن نستخدم libraries, children's libraries، نستخدم structured language. والمثال الكامل على اللغة المصطنعة هو رمز خطة التصنيف، فبدلا من مصطلحات اللغة الطبيعية معالجة الألومنيوم بالحرارة، أو الشكل الأكثر صياغة رسمية aluminium - heat treatment، نستخدم ٠٤٦٦٩٧١٠٤.

إن استخدام لغة مصطنعة يمكننا من استخدام تكشف المفاهيم concept analysis وليس تكشف المصطلحات term indexing. فى تكشف المصطلحات نعتد على الكلمات التى يستخدمها كل مؤلف لكى تعطينا أوصاف الموضوع الذى نحتاجه، والذى نأمل أننا حينما نحاول أن نجد المعلومات سوف نضاهي الكلمات المختلفة التى استخدمها مؤلفون مختلفون؛ وهذه هى مضاهاة الكلمات التى نستخدمها فى بحث الكلمة المفتاح بواسطة الحاسب. أما فى كشف المفاهيم، فإننا نحاول تكوين وصف موحد (قياسى) لكل مفهوم وأن نستخدم ذلك الوصف فى كل مرة يكون فيها مناسبا، سواء استخدمه المؤلف أم لا. وحينما نبحت ملفاتنا، فإننا نستخدم الوصف الموحد مرة أخرى، وينبغى أن نكون قادرين على مضاهاة هذه بطريقة أكثر أطرادا أو أكثر اعتمادية. واللغة الطبيعية مرنة جدا، وبصفة خاصة فى لغة عالية التطور مثل الانجليزية وهذه ميزة هائلة للمؤلفين، الذين يمكنهم أن يراجعوا مصطلحاتهم للمحافظة على اهتمام القارئ، ولكنها عائق أمام المكشف، الذى يهتم بالأفكار التى تحملها الكلمات أكثر مما يهتم بلطائف الأسلوب الأدبى.

فإذا استخدمنا الكلمات لأوصاف موضوعنا، فإن الطريقة الوحيدة التى يمكن أن نستخدمها فى ترتيبها هى الترتيب الهجائى؛ فإذا حاولنا أن نرتبها بأية طريقة أخرى فلن نكون قادرين على الوصول إليها حينما نريدها. ولكن هناك مناسبات كثيرة يكون الترتيب الهجائى غير مفيد بصفة خاصة؛ مثال ذلك:

إذا أردنا أن نتبع تقدم السكة الحديدية، فإن قائمة هجائية بالمراحل سوف تكون أقل فائدة بكثير من تسلسل زمني. وإن لغة اصطناعية مثل الرمز في خطة التصنيف تمكننا من ترتيب المفاهيم بأى طريقة نرغبها، وتحسن فرصنا فى الوصول إلى المعلومات حينما نبحث عنها، مرتبة بالطريقة التي نريدها عليها. وتنطوى لغة التكشيف المعين على عمل من جانب المشكف ومن جانب الباحث؛ وبالمقابل، يجب أن تساعدنا فى تحسين كل من الاستدعاء والصلة بالصورة التي نحتاجها.

الصلة و «المضمون»

افترضنا حتى الآن أننا يمكن أن نقرر ما الذى تدور حوله الوثائق (المضمون)، وما نوع المعلومات التي يبحث عنها القراء، حتى يتسنى لنا أن نصدر أحكاما على الوثائق أيها يتصل بموضوع السؤال. وعند مناقشة الاسترجاع فى النظم المبنية على الحاسب افترضنا أننا عن طريق مضاهاة كلمات القارئ كما جاءت فى سؤاله مع تلك التي استخدمتها الوثيقة، ربما باستخدام مختلف الوسائل لتعديل السؤال الأصلي، يمكننا أن نحدد مكان الوثائق التي تلبى احتياجات القارئ. وفى النظم اليدوية، أشرنا إلى الطرق التي يمكن أن نبني بها الرؤوس ونتداولها والتي تمكننا من مضاهاة احتياجات القراء. كيف يمكن أن نوزع هذه الرؤوس بصورة مناسبة؟ كيف يقرر القراء أن وثيقة ما متصلة أم لا؟ ما معايير الحكم على أداء المكشف / الباحث أو الحاسب؟ رأينا أن الصلة حكم ذاتي، ولذلك ينبغي ألا نتوقع أية إجابة سهلة عن هذا السؤال، ولكن يمكن أن ندرس به الأفكار التي طرحت.

فى بداية الفصل ٣ تناولنا بعض الطرق التي يمكن بواسطتها أن نحدد موضوع الوثيقة؛ ومن هذه يمكن أن نتفق مع هتشنس Hutchins^(١٢) حينما نقل عن دليل بريسييس^(١٣). إلى درجة أن «المكشف» بعد أن يفحص الوثيقة، يكون قد كون فى ذهنه شيئا عن تسلسل ذى معنى للكلمات التي تلخص المادة الموضوعية، وهذا وصف معقول لأسلوب عملي، ولكنه لا يأخذ حقيقة إلى أبعد من ذلك

كثيرا. يقترح هتشنس طريقة يمكننا فيها أن نقرر ما الذى تدور حوله الوثيقة، معتمدا على البنى اللغوية linguistic. وفى أية قطعة من نص، سوف تكون هناك معلومات معينة يعتبرها المؤلف «مسلمة» أى يفترض أن القارئ يعرفها أصلا. وسوف توجد معلومات أخرى يفترض أنها «جديدة»، أى لا يعرفها القارئ أصلا. ولذلك فعند دراسة وثيقة ما لأغراض التكشيف، يمكن أن نحدد المعلومات المسلمة والمعلومات الجديدة باعتبارهما ربطتين مستقلتين. كذلك يمكن تقسيم القراء إلى فئتين : غير المتمرس (المبتدىء) novice، الذى يعرف القليل عن الموضوع والذى يرغب فى معرفة شئ عنه، والخبير، الذى يعرف الكثير عن الموضوع ولكنه يرغب فى توسيع معرفته إذا كان ذلك ممكنا. (كما ذكرنا سابقا أن العالم الحقيقى رمادى، لا أسود ولا أبيض، ولكن التعميم يعكس إلى حد كبير توفير الخدمات المكتبية: المكتبة العامة توفر حاجات الشخص العام إلى حد كبير، والمكتبة المتخصصة حاجات المتخصص). والصورة المثالية هى أن المعلومات «المسلمة» فى وثيقة ما ينبغى أن تضاهاى أى تتوافق مع معرفة القارئ وقتها، فى حين أن المعلومات «الجديدة» ينبغى أن تتوافق مع الحاجة الجارية. ولما كانت عملية الاتصال ليست بهذه الدرجة من الدقة أو الإحكام، فلن نكون قادرين على مضاهاة هذا النموذج بالضبط، ولكنه هدف يمكن أن نسعى إليه. ولذلك فإننا نحتاج إلى أن نكون واضحين : أى نوع من القراء نسعى إلى خدمته.

ولاستطيع الحاسب أن يميز بين المعلومات «المسلمة» و «الجديدة». ولاينتظر أن تتجاوز «المسلمة» حدود معرفة الخبير، فى حين أننا نرحب بأية معلومات «جديدة» ومن ثم فمن المنتظر أن الخبير سوف يرضى بنتائج بحث الحاسب. ومن ناحية أخرى، وكما يحتاج هتشنس، فإن المبتدئين يحسن أن يستخدموا فقط المعلومات «المسلمة» كأساس للتكشيف، حيث أن هذه سوف تقود إلى نوع الوثائق التى سوف تلبى على أفضل صورة حاجتهم لتوسيع معلوماتهم الموجودة.

ومن الناحية العملية، نجد أن المكشفين في المكتبات الشاملة يعتمدون حقيقة وإلى حد كبير على تلك الأجزاء من الوثيقة التي يضع المؤلف فيها المعرفة المفترضة التي تكون أساس العمل؛ التصدير، المقدمة، قائمة المحتويات وما إلى ذلك. وبالمقارنة، ينزع المكشفون في المكتبات المتخصصة إلى استخدام التشفير لأي عدد مكن من الأفكار themes في الوثيقة. ويشير هتشنس أن خدمات الاستخلاص توفر تلخيصا جاهزا للوثيقة، يشتمل على كل من المعلومات المسلمة والجديدة، وهكذا تواجه احتياجات الخبير الذي من أجله أعدت هذه المستلخصات. وهذا يتفق مع نتائج تجارب التقييم مثل كرانفيلد^(١٤)، التي أوضحت أن النظم المبنية على المستلخصات تعطى أفضل النتائج. وقد حاجج لانكستر بقوة بأن الاستخلاص والتشفير هما وجهان لعملة واحدة^(١٥).

وفي المؤتمر نفسه حاجج سويفت Swift^(١٦) بأن معيار المحتوى التقريبي والذي يكون مفيدا في العلوم والتقانة قد لا يكون مفيدا بالدرجة نفسها في العلوم الاجتماعية. وقد اتفق كل من إليس Ellis وبيكنز Pichins على هذه النقطة حديثا جدا^(١٧). وتقرح إليس أن العلماء والاجتماعيين يتبعون نماذج للبحث عن المعلومات تختلف عن النماذج أو الأطر التي يتبعها العلماء والتقانيون بينما يفترض بيكنز أن بنية قاعدة البيانات ينبغي أن تكيف بحيث تكون أكثر قربا لاحتياجات المستفيدين. مثال ذلك: التربية لها مجالات أربعة كبرى: الموضوعات؛ المحتوى؛ المنطقة؛ والشكل. الموضوعات (المتعلمون) يمكن تمييزهم بواسطة العمر أو المعهد، المحتوى (المادة الموضوعية) يمكن تمييزها بواسطة الواصفات، التي قد تكون مهمة أو أقل أهمية؛ المنطقة (المكان) قد يكون مكان الأصل أو مكانا كتب عنه؛ والشكل (نوع الوثيقة) مهم من حيث أنه يضاهاى نوع الوثيقة مع القارىء. وفي قاعدة بيانات إريك، تغطى كل هذه، ولكن في مجال الواصفة نفسه، مع استثناء المكان، الذي يعامل كمعرف -identi-fier وقد يكون أكثر فائدة للمستفيد النموذجي أن تعالج هذه المجالات بصورة

مستقلة. ويقترح بيكنز أن تصميم قاعدة البيانات لا يزال مقيدا بالحدود القديمة (لتقانة المعلومات) والتي لم يعد لها أهمية.

ويعرف لانكستر^(١٩). العوامل نفسها ولكن بطريقة مختلفة اختلافا يسيرا. يمكن اعتبار أن «المحتوى التقريبي» داخلي intrinsic - موضوع الوثيقة - أو خارجي extrinsic - الاعتبارات الخارجية مثل الغرض الذي من أجله تطلب الوثائق، أو من الذي يطلبها. وإن أوصاف الوثائق، والأسئلة، ولو غاريمتات (خوارزميات) المضاهاة كلها مستقلة. ويجب ألا تدرس في فراغ. كذلك يجب أن نميز بين وسائل (طرق) لغة التكشيف مثل الأدوار، ووسائل التكشيف / البحث مثل الوزن. فالأدوار هي أساسا جزء من عملية المدخلات، التي يحاول المكشف فيها أن يضيف إلى خصوصية التكشيف؛ وهي هكذا مربوطة بنظرة المكشف إلى الصلة. أما الوزن فهو في العادة جزء من عملية البحث، وهو يعكس تقدير المستفيد للصلة. وقد اقترح (اقترحت) مارون، كما ناقشنا في الفصل ٥ أنه من خلال جمع وزن كل خلية wiz في مصفوفة المصطلح - الوثيقة، يمكننا أن نصل إلى مقياس للمحتوى التقريبي للوثيقة باعتباره متجه vector العمود الموزون لتلك الوثيقة^(٢٠).

ويجب كذلك أن ندرس الطرق المختلفة التي يحدس فيها التقدم في الحقول الموضوعية المختلفة. فالعلم يتقدم أساسا بواسطة الإجماع (الاصطلاح) con-census^(٢١)، في حين يبنى التقدم في العلوم الاجتماعية والانسانيات في العادة على عملية جدلية، ولا يكون الإجماع (الاصطلاح) هو الهدف الأول.

وكما رأينا، فإن الصلة هي حكم إجماعي، والتي قد لا تكون مناسبة في مجالات أخرى غير العلم والتقانة.

الاطراد

رأينا أن الأطراد عامل مهم جدا في الاسترجاع الناجح للمعلومات. ونحن نحتاج إلى الاطراد في اختيارنا للمصطحات للتعبير عن المفاهيم المختلفة،

وكذلك فى الطريفة التى نعبر بها عن الموضوعات المركبة، ولكن هناك عددا من العوامل التى تعمل ضد هذا. بعض هذه العوامل يعود إلى المكشفين، وحيث أننا جميعا أفراد، فإن لكل منا جميعا نموذجا خاصا أو إطارا خاصا للرجوع، وخلفية تعليمية خاصة بنا، وتجربتنا فى الحياة، وقد يكون من الصعب بالنسبة لنا أن نكون مطردين الواحد منا مع الآخر فى الطريفة التى نفسر بها ما نحاول تكشيفه. وآراؤنا الخاصة تتغير مع الوقت؛ وكما أشار هرقليطس Heraclitus فإن من غير الممكن أن نخطو مرتين فى النهر نفسه. هناك اتجاه موروث (ملازم) نحو عدم الاطراد، أى التضارب. كيف نواجه هذا؟

فى هذا الفصل والذى يسبقه درسنا الطرق المختلفة التى يمكن من خلالها أن نحاول أن نحقق الاطراد - فإذا كانت لدينا قواعد واضحة لاختيار المصطلحات، مقرر فى لغة مقيدة، فنحن أقرب إلى الاطراد منا فى حالة اعتمادنا على الكلمات التى يستخدمها المؤلفون فقط. وإذا كانت لدينا قواعد واضحة عن كيفية معالجة الموضوعات المركبة، فنحن مرة ثانية أقرب إلى الاطراد منا فى حالة عدمها. وإن الوقت الذى ينفق فى تكشيف وثيقة ما سوف يؤثر على طريفة تكشيفنا لها. إذا كان الوقت أقصر مما ينبغى، فسوف لا يتسع إلا للتصفح السريع دون أن نستوعب بصورة شافية المضمون الكامل للوثيقة، وإذا كان أكثر من اللازم، فقد نبدأ فى أن نحلم بمظاهر غير موجودة سوف لا تكون مفيدة لهؤلاء الذين سوف يسترجعون الوثيقة فى النهاية. وإن الخبرة بكل من التكشيف والاسترجاع مهمة؛ وقد ذهبت الأيام التى كان المفهرسون يستطيعون أن يجلسوا فى الغرفة الخلفية دون أن يكون عليهم حتى أن يقابلوا الناس الذين يفترض أنهم يساعدونهم، على الأقل فى معظم الأوضاع. ومع ذلك، فمن المنتظر أننا سوف يتعين علينا أن نقبل أن التضارب وعدم الاطراد هو المجازفة أو المخاطرة التى يجب أن نعيش معها، وأن نكون مستعدين للتغلب عليها، بينما نحاول أن نبقىها فى الحدود الدنيا^(٢٢). ولسوء الحظ، فإنه يبدو أن روابط النصوص الفائقة هي عرضة للتضارب مثل أى نوع آخر من التكشيف^(٢٣)، وأن السهولة التى يمكن بها خلق

علامات الكتب في World Wide Web تقترح أنها أيضا يمكن أن تكون مصدرا
خصبا للمشاكل!

ملخص

رأينا أن جانبا جوهريا من استرجاع المعلومات ينفذه الحاسب الآن، بالاعتماد على اللغة الطبيعية للوثائق في قاعدة البيانات. وقد رأينا أيضا أن ذلك لا يكون مفيدا لأغراض معينة - ترتيب الرفوف، الكشافات المطبوعة، وأنا نحتاج إلى تعيين مصطلحات الكشاف للوثائق لكي تحل محل النص؛ هذا التكشيف قد يكون ذا قيمة أيضا في بحث الحاسب لكي يكمل النص في حالات معينة. وإن مجموعة من المصطلحات، مع العلاقات بينها، يمكن أن تكون لغة للتكشيف، مع المفردات وربما أيضا التراكيب. وقد تختلف حاجات القراء من علم إلى آخر، وهذا يجعل من الصعب تأسيس مقياس متفق عليه بصفة عامة «للمضمون». والتضارب أو عدم الاطراد، سواء بالنسبة للمكشيف نفسه أو فيما بين المصنفين، هو سمة دائمة لكل نظم استرجاع المعلومات. ومع ذلك فإن الغرض من تلك الأنظمة يظل هو مضاهاة حاجات القراء مع مخرجات المؤلفين بأقصى قدر ممكن من الفاعلية والفائدة.

المراجع

- 1 Farradane, J. E. L. et al., *Report on research into information retrieval by relational indexing: part 1: methodology*. London, City University, 1966.
- 2 Gardin, J. C., *Syntol*, Rutgers, The State University School of Library Science, 1965. (Rutgers series on systems for the intellectual organization of information.)
- 3 Keen, E. M. and Digger, J., *Report of an information science index languages test*. Aberystwyth, College of Librarianship Wales, 1972. 2v.
- 4 Chan, L. M., 'Interindexer consistency in subject cataloging', *Information technology and libraries*, 8, 1989, 349-58.
- 5 O'Brien, A., 'Online catalogs: enhancements and developments', *Annual review of information science and technology*, 29, 1994, 219-42.

- 6 Keen, E. M., 'On the generation and searching of entries in printed subject indexes', *Journal of documentation*, **33** (1), 1977, 15-45.
- Keen, E. M., 'On the processing of printed subject index entries during searching', *Journal of documentation*, **33** (4), 1977, 266-76.
- Keen, E. M., *On the performance of nine printed index types: a selective report of EPSILON*. Aberystwyth, College of Librarianship Wales, 1978.
- 7 Foskett, D. J., 'Two notes on indexing techniques'. *Journal of documentation*, **18** (4), 1962, 188-92.
- 8 The phase relationships demonstrated here are those identified by *BNB*.
- 9 Chomsky, N., quoted by Austin in the *PRECIS Manual*, 1st edn, Chapter 16.
- 10 Berman, S. *Prejudices and antipathies*. Metuchen, NJ, Scarecrow Press, 1971.
- Berman, S., *Worth noting: editorials, letters, essays, an interview and bibliography, with a foreword by Bill Katz*. Jefferson, NC, McFarlane, c1988.
- Much of Berman's writing is concerned with this aspect of indexing languages.
- Fina, M., 'The role of subject headings in access to information: the experience of one Spanish-speaking patron', *Cataloging and classification quarterly*, **17** (1/2), 1993, 267-74. (The WASP bias causes problems for minority groups.)
- Foskett, A. C., 'Misogynists all: a study in critical classification', *Library resources and technical services*, **15** (4), Spring 1971, 117-21.
- Foskett, A. C., 'Better dead than read: further studies in critical classification', *Library resources and technical services*, **28** (4), Oct/Dec. 1984, 346-59.
- Rogers, M. N., 'Are we on equal terms yet? Subject headings concerning women in LCSH 1975-1991' *Library resources and technical services*, **37** (2), 1993, 181-96.
- Wilson, A. 'The hierarchy of belief: ideological tendentiousness in universal classification', in *Classification research for knowledge representation and organization: proceedings of the 5th International study conference on classification research, Toronto, Canada, June 24-28 1991*, Williamson, N. J. and Hudon, M. (eds.), Elsevier, 1992. (FID 698), 389-97.
- 11 Sheriff, S., 'Library of Congress insults Islam', *Library Association record*, **88** (8), August 1986, 371.
- 12 Hutchins, W. J., 'The concept of 'aboutness' in subject indexing', *Aslib proceedings*, **30** (5), 1978, 172-81. Coordinate Indexing Group Colloquium on aboutness, April 1977.
- 13 Austin, D., *PRECIS: a manual of concept analysis and subject indexing*, London, British National Bibliography, 1974, 4.
- 14 Cleverdon, C. W., Keen, E. M. and Mills, J., *Factors determining the performance of indexing systems*. Cranfield, Aslib-Cranfield Research Project, 1966. 2v. in 3.
- 15 Lancaster, F. W., *Indexing and abstracting in theory and practice*, London, Library Association, 1991.
- 16 Swift, D. F., Winn, V. and Bramer, D., "'Aboutness" as a strategy for retrieval in the social sciences'. *Aslib proceedings*, **30** (5), 1978, 182-7. Coordinate Indexing Group Colloquium on aboutness. April 1977.
- 17 Ellis, D., 'A behavioural approach to information retrieval system design',

- Journal of documentation*, 45 (3), 1989, 171-212.
- 18 Pickens, K., 'The relationship of bibliographic database design to the structure of information: a case study in education', *Journal of documentation*, 50 (1), March 1994, 36-44.
 - 19 Lancaster, F. W., Elliken, C. and Connell, T. H., 'Subject analysis', *Annual review of information science and technology*, 24, 1989, 35-84.
 - 20 Maron, M. E., 'On indexing, retrieval and the meaning of about', *Journal of the American Society for Information Science*, 28 (1), 1977, 38-43.
Stone, A. J., 'That elusive quality of "aboutness": the year's work in subject analysis, 1992', *Library resources and technical services*, 37 (3), 1993, 277-98.
 - 21 Ziman, J., *Public knowledge: the social dimension of science*, Cambridge, Cambridge University Press, 1969.
 - 22 Leonard, L. E., *Inter-indexer consistency studies 1954-1975: a review of the literature and a summary of study results*, Urbana, IL. University of Illinois Graduate School of Library Science, Occasional papers No 131.
 - 23 Ellis, D., Turner-Hines, J. and Willett, P., 'On the creation of hypertext links in full-text documents: measurements of interlinker consistency', *Journal of documentation*, 50 (2), 1994, 67-98.
Langford, D. and Brown, P., 'Creating hypertext comments: is it worth the effort?', *Aslib proceedings*, 45 (4), 1993, 91-5. Suggests the use of commentary (hidden text) to explain reasons for links.