التوثيق الآلى ودوره فى اختزان واسترجاع المعلومات

إعداد
أ.د. شعبان عبد العزيز خليفة
أستاذ ورئيس قسم المكتبات والوثائق والمعلومات
كلية الآداب ـ جامعة القاهرة

يوصف العصر الذى نعيش فيه فى نهاية القرن العشرين بأنه عصر المعلومات حيث تصدر فيه كميات هائلة من مصادر المعلومات وحيث لايمكن اتخاذ أى قرار لا على المستوى الرسمى ولا على المستوى الشخصى ولا على مستوى البحث العلمى إلا بالاستناد إلى المعلومات الدقيقة والصحيحة وفى الوقت المناسب. يصدر فى العالم اليوم فى كل سنة نحو مليون كتاب ونحو نصف مليون دورية ـ منها مائة وستون ألف دورية متخصصة ـ ونحو ٢ مليون مصغر فيلمى و ٢ مليون مادة سمعية بصرية ومائة ألف ملف بيانات آلية ونحو خمسين ألف قرص ليزر، ناهيك عن ملايين التقارير الفنية والبحثية ومليارات الوثائق الأرشيفية.

هذه المصادر تحمل كميات فلكية من المعلومات وعلى سبيل المثال يقدر المحافظون عدد المقالات التى تصدر فى الدوريات فى كل يوم بنحو مائة ألف مقالة.

هذا الإنتاج الفكرى الضخم يصدر بكثافات مختلفة فى أنحاء متفرقة من العالم، وبلغات شتى عديدة، ولا يستطيع أى عالم مهما دق تخصصه أن يحيط بكل أو جل ما نشر فى تخصصه إن لم يكن بسبب الحواجز الجغرافية فبسبب الحواجز اللغوية والحواجز الشخصية من وقت وجهد ومال.

وإذا كان التقدم فيما يقول ابن خلدون (١٣٣٢ - ١٤٠٦م) هو أن نبدأ من حيث انتهى الآخرون، فإننا لن نعرف أين انتهى الآخرون إلا إذا عرفنا ماذا أنتج الآخرون.

من هنا كانت أهمية ذلك الاختراع الفذ «التوثيق» وبصرف النظر عن تقلب مفهوم التوثيق وتطوره فى الزمان والمكان، إلا أننا فى هذا المقام سنضع له التعريف الإجرائى الآتى:

«التوثيق هو حصر وتسجيل ووصف الإنتاج الفكرى وتحليل محتوياته والتعريف به بقصد تيسير الإفادة منه».

والحقيقة أن التوثيق بهذا المعنى بدأ منذ بدأ ظهور الإنتاج الفكرى نفسه، ولكنه ظل يتم بطريقة يدوية طالما بقى كم الإنتاج الفكرى ونوعه فى حدود السيطرة اليدوية، بيد أنه عندما حدث الانفجار الفكرى ووقعت ثورة المعلومات فى نهاية قرننا العشرين على النحو الذى أشرت إلى جانب منه، لم تعد السيطرة اليدوية على الإنتاج الفكرى ممكنة أو ذات جدوى كبيرة وتطلب الأمر استغلال إمكانيات تكنولوجيا المعلومات فى هذا الصدد.

وجاء الحاسب الآلى فى موعده تماماً مع تفجر المعلومات وثورتها وكبر معها وبها وكبرت به ومعه.

فبعد عقد واحد من ظهور الحاسب الآلى وبعد أن تعددت وظائفه خارج عمليات الحساب، استخدم فى عمليات التوثيق واختزان واسترجاع البيانات الببليوجرافية. وكان استخدام الحاسب الآلى فى هذا الصدد قد بدأ سنة ١٩٥٦ حين نظم أول مؤتمر عن الترجمة الآلية فى شهر أكتوبر من ذلك العام وكانت أول محاولة تطبيقية عملية فى التوثيق الآلى على يد هانز بيترلون فى شركة آى. بى. إم سنة ١٩٥٨. وقد تطورت عملية التوثيق الآلى بعد تلك المحاولات الفجة إلى عملية أكثر تعقيداً كما سنرى فيما بعد.

والتوثيق عندما كان يتم يدوياً على يد البشر كان عبارة عن عملية تحليل عقلانى للمادة العلمية يبنى على فهم معنى النص بهدف إعادة ترتيب المفاهيم الموجودة فيه وبشكل يسهل البحث فيه، وينتج عن ذلك تركيب متواز للنص ينظم تنظيماً مختلفاً.

أما التوثيق الآلى فإنه يترك النص على حاله ويضع استراتيجيات للبحث فيه؛ هذه الاستراتيجيات تصمم أساساً بهدف إيجاد الكلمات التى يستخدمها الباحثون فى بحثهم وتساؤلاتهم ومقابلة هذه الكلمات مع نظيراتها الموجودة فى نصوص اللغة الطبيعية. ولكننا إزاء هذه الاستراتيجيات الآلية يجب أن نؤكد على:

١- أن ما تقدمه قواعد البيانات الببليوجرافية ليس إلا مصفوفات من استراتيجيات البحث المختلفة.

٢- أن مضاهاة الكلمات (أو أجزائها) والجمل لا يعنى بالضرورة تطابق المفاهيم مع الكلمات والجمل التى يريدها الباحث، ومن ثم فإن هذه المضاهاة قد تفشل فى استرجاع الوثائق المناسبة. ومن ثم فليس صحيحاً ما يقال عن التوثيق الآلى من أنه «دفع خلفى للسيارة».

ومن الواضح بذاته أنه لكى نبحث فى نص كامل باللغة الطبيعية بطرق آلية، فإن هذا النص لابد وأن يكون مختزناً ومتاحاً فى شكل مقروء آلياً. وهذا الأمر يحد من أنواع النصوص التى يمكن توثيقها آلياً ويقصرها على المقالات والتقارير وغيرها من الوثائق القصيرة وبما يعنى استبعاد معظم الكتب. وحيث دلت التجارب على أن كشافات الكتب الآلية لا يمكن أن تنافس تلك المعدة يدوياً فى نهاية الكتاب.

إن للتوثيق اليدوى والآلى نفس الهدف: وهو جعل الوثائق سهلة الاسترجاع من جانب هؤلاء الذين يحتاجونها أو يرغبون فيها. والتوثيق الآلى يضع عبء الاسترجاع على المستفيد النهائى الذى عليه أن يجد الكلمات التى تم استخدامها فى النص، وإلا لن يحالفه الحظ فى الاسترجاع.

ولا يوجد من بين أنظمة التوثيق الآلى الحالية ما يقدم للباحث المترادفات المساوية للمصطلحات المستخدمة فى النص ولا أى منها يمكنه عرض العلاقات القائمة بين المفاهيم التى تتشتت فى النص ولا المفاهيم التى وردت ضمنية فى النص، أى تلك التى تفهم فقط من السياق ومن بين السطور. ولكى تساعد المستفيد النهائى وتعوضه بعض الشىء عن نقائص التوثيق الآلى تقوم بعض قواعد المعلومات كما سنرى فيما بعد بإمداده بمكانز مباشرة أو مطبوعة تقدم له المترادفات والمصطلحات ذات العلاقة. ولكن يجب أن يفهم أن تلك المكانز ليست جزءاً من أى نظام للتوثيق الآلى ولكنها ضمن مكونات نظام الاسترجاع، بل وأكثر من هذا فإن تلك المكانز التى تعرض شبكة من المصطلحات ذات الصلة الموضوعية لا يمكنها أن تستوعب معظم المصطلحات الضرورية للاسترجاع الناجح. وهى من جهة ثانية أدوات شديدة التعقيد قد لا يفيد منها إلا المكشفون المحترفون وليس

المستفيدون الذين قد لا يعرفون كيف يستخدمونها بل لا يكلفون أنفسهم عناء النظر فى المصطلحات التى تضمها. ونريد أن نؤكد على أنه لا يوجد مكنز واحد يستطيع أن يساعد فى استرجاع المفاهيم الضمنية فى النص.

إن التوثيق اليدوى على الجانب الآخر لا يقدم فقط المصطلحات الدالة على المفاهيم الموجودة فى النص بالوثيقة، ولكن أيضاً مرادفاتها وعلاقاتها بالمفاهيم الأخرى فى بناء منطقى ويمكن استيعابه. إنه يفعل ذلك عن طريق رؤوس الموضوعات متعددة المستويات، وعن طريق الإحالات المزدوجة وغير ذلك من الوسائل التى تعتبر جزءاً من البنية العامة للكشاف والتى لا تتطلب استخدامها منفصلة مستقلة من جانب المستفيد.

إن جميع نظم التوثيق الآلى تبنى على ظهور وتداعى الكلمات فى النصوص وعند هذه النقطة فلابد وأن نتوقف أمام: ماذا يجعل الكلمة صالحة لأغراض التوثيق الآلى؛ من حيث المبدأ ومن الناحية النظرية العامة؟ إنها أية وحدة مكتوبة فى النص مفصولة عن الوحدات الأخرى بمسافات أو علامات ترقيم وهناك عدة استثناءات من هذه القاعدة، وخاصة فيما يتعلق بعلامات الترقيم مثل الشرطة والمسافة والفاصلة والأقواس التى قد تعتبر إما جزءاً من كلمة أو تهمل استناداً إلى نظام التوثيق الآلى ولوغاريتم الفرز فى الحاسب المستخدم.

نظم الاسترجاع الآلى:

هناك عدد من الأدوات التى يمكن عن طريقها استرجاع المادة العلمية المختزنة فى الحاسب من بينها:

البحث بالكلمات الدالة. وهى الطريقة القديمة فى التوثيق الآلى. وكانت فى الأصل تستخدم لاسترجاع عناوين الوثائق فقط. وكانت تفترض أن عناوين المقالات والتقارير تعبر ببساطة عن موضوعاتها. وهو افتراض غير صحيح حتى بالنسبة لأكثر المقالات تخصصاً.

وتعتمد كل نظم التوثيق الآلى على «**قوائم إيقاف**». الكلمات التى يعتقد أنها غير هامة أو غير ضرورية لأغراض الاسترجاع مثل أدوات التعريف والتنكير والضمائر وحروف الجر وأدوات الوصل والظرف والتى تتردد كثيراً فى النصوص والتى ليس لها معنى فى حد ذاتها. وبعض قوائم الإيقاف قد يحمل مئات من الكلمات من الضمائر والصفات إذ تعتبر لا قيمة لها فى السياق. وكلما كانت قاعدة البيانات كبيرة كلما كانت قائمة الإيقاف قصيرة ومن مشاكلها أن وقف أداة التعريف فى العربية «أل» قد يتسبب فى عدم استرجاع كلمات «الله»، ألفونس وغيرهما. وفى الإنجليزية قد يتسبب إيقاف أداة التنكير A فى عدم استرجاع Vitamin A وهكذا.

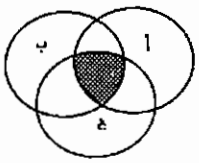
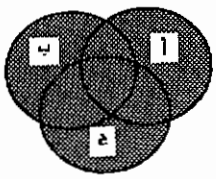
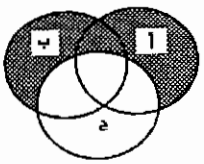
وتسمح **أوامر الربط والتقريب** للمستفيدين بتحديد أن مصطلحين يجب أن يتقاربا من بعضهما حتى يسترجعا معاً أو أنهما يجب أن يقعا على مسافة محددة (عدد من الكلمات) من بعضهما. وهذه الطريقة ليست جزءاً من التوثيق الآلى ولكنها ملمح من ملامح نظم الاسترجاع، وتساعد فى بحث النصوص بطريقة آلية.

والتجذير هو تنقية الكلمات الدالة وإرجاعها إلى أصولها بحيث تسمح للباحث أن يحدد فقط جذور الكلمات التى قد يكون لها بادئة أو لاحقة.

الجدع أو البتر وهو قريب الصلة من التجذير ويهدف إلى حذف بعض حروف من نهاية الكلمات فى حالة الجدع الأيمن أو من بداية الكلمات فى حالة الجدع الأيسر، وهناك كذلك جدع الغرز

الذى يضيف حرفًا فى وسط الكلمة، والذى يستخدم أساساً للبحث عن المركبات الكيميائية ذات الصلة فى الوسط ولكنها غير محددة.

المنطق البوليانى. وقد سمى باسم مخترعه جورج بول فى القرن التاسع عشر والذى يساعد المستفيد على تشكيل بحث عن علاقات عدة مصطلحات بثلاث طرق مختلفة، والذى يكشف عنه الرسم المرفق. وبسبب الاستخدام غير التقليدى للكلمات: **و** AND، **أو** OR ، **ليس** NOT والتكوينات المختلفة التى تستخدم فيها هذه المفاتيح فى حالة البحث المعقد ـ مثال ذلك (أ **و** ب) **أو** ج **وليس** د ـ بسبب ذلك فإن كثيراً من المستفيدين لا يستطيعون تطبيق المنطق البوليانى تطبيقاً دقيقاً. وفى إحدى قواعد البيانات الشهيرة فى سنة ١٩٩٧م أجريت ٢٤٥٥٢١ عملية بحث، ٢٧ ٪ منها لم تسترجع أية مداخل على الإطلاق، ٥٧ ٪ من بينها كان بسبب المنطلق البوليانى.

المفتاح	نوع البحث	تصوير العلاقة	شرح البحث
و	**الضرب**		الناتج المنطقى: أ × ب × ج عبر عنه بـ أ، ب، ج وحيث يجب أن يكون كل من أ، ب، ج فى الوثيقة. مثال ذلك التكشيف والميكنة والكتب .
أو	**الجمع**		حاصل المنطق: أ + ب + ج يعبر عنها بـ أ أو ب أو ج. أى من المصطلحات الثلاثة فى الوثيقة. مثال ذلك التكشيف أو الميكنة أو الكتب.
ليس	**الطرح**		الفرق المنطقى أ × ب ـ ج ويعبر عنه أ و ب وليس ج، فقط أ و ب وليس ج مما قد يوجد فى الوثائق. مثال ذلك: التكشيف والميكنة وليس الكتب.

وزن المصطلح يعتبر واحداً من أكثر نظم التوثيق الآلى تعقيداً، وقد تطور منذ أوائل الستينيات بقصد التغلب على مشاكل البحث بالكلمات الدالة. ويقوم ذلك النظام على استخدام طرق إحصائية واحتمالية مختلفة لترتيب المصطلحات طبقا لمعدلات ترددها فى مجموعة الوثائق حتى نجد الوثائق التى تناسب السؤال. بيد أن هذا النظام يصلح للتطبيق على مجموعات صغيرة نسبياً من الوثائق ويفشل فى الحصول على نتائج أفضل من النظم الأخرى فى حالة قواعد البيانات الكبيرة التى تضم ملايين التسجيلات.

الطرق اللغوية مثل مضاهاة الجمل، وإعراب الكلمات والتراكيب اللغوية فى النص، بل وحتى تحليل بعض المعانى بمساعدة المعاجم والمكانز، ولقد جرت بعض المحاولات التجريبية فى هذا الصدد ولكن

قيمتها العملية ماتزال حتى الآن مشكوكاً فيها. وغالباً ما يفشل النظام فى التعرف على الموصوف كما يفشل فى التفريق بين الاستخدام المجازى والاستخدام الحقيقى للكلمات.

ورغم أن الطرق اللغوية قد تصادف بعض النجاح فى المجالات المتخصصة نظراً للثبات النسبى والتوحيد فى المصطلحات المستخدمة، إلا أنها لاتستخدم فى نظم الاسترجاع الواسعة النطاق أو التجارية فى الوقت الحاضر.

ويعتقد الخبراء أن مزيجاً من النظم المتعددة فى التوثيق الآلى قد يكون علاجاً شافياً لمشكلات البحث فى مجموعات الوثائق الكبيرة، ولكن هذا المزيج لم يتحقق بعد.

❊ ❊ ❊

التوثيق بالمصطلحات الحرة :

لقد بدأت عملية التوثيق الآلى بما عرف بالمصطلحات الحرة أو غير المقيدة. وهذه المصطلحات الحرة مرت بعدد من المراحل:

١- الكلمات الدالة فى السياق KWIC: ويقصد بالسياق هنا أساساً عنوان العمل حيث تؤخذ الكلمات المفتاحية أو ذات الدلالة الاصطلاحية وتعتبر مدخلاً إلى النص. وعيوب ذلك الأمر هو تعدد المصطلحات المترادفة الدالة على مفهوم واحد ومن ثم تشتت المادة العلمية تحت العديد من المصطلحات، وأكثر من هذا ربما تختلف المصطلحات التى يستخدمها المستفيدون فى استرجاع الوثائق. ففى بحث مستفيض عن الطريقة التى يسمى بها الناس الأشياء عند الاسترجاع، وجد أن شخصين اتفقا فى نفس المصطلح فى ٢٠٪ فقط من الحالات وأن ٨٠٪ من الحالات استخدمت فى المتوسط ١٥ مصطلحاً فى استرجاع نفس الشيء.

وقد اتفق الباحثون على أنه ليس هناك مصطلح واحد صالح لاسترجاع معظم الأشياء وأن فكرة المصطلح «الواضح» «القائم بذاته» أو «الطبيعى» هى مجرد خرافة.

٢- الكلمات الدالة خارج السياق KWOC: أى أن نعتمد على المصطلحات الأكثر استخداماً من جانب المؤلفين والمستفيدين على السواء وبصرف النظر عن المصطلحات الواردة فى الوثائق قيد التكشيف. ولكن المشكلة هنا هى أن المصطلحات المستقاة من خارج السياق تخضع هى الأخرى لمعايير شخصية، ومن ثم لا يكون هناك التوحيد من جانب الموثقين أنفسهم ولا يكون هناك سهولة الإدراك من جانب المستفيدين أيضاً على نحو ما صادفناه فى المرحلة السابقة.

٣- الكلمات الدالة من خارج وداخل السياق KWAC: وهى مزيج من الطريقتين السابقتين فى محاولة للتوحيد والتقنين ولكنها ماتزال تخضع للمعايير الشخصية البحتة التى تختلف من موثق إلى آخر فى الوقت الواحد وللموثق الواحد فى أوقات مختلفة من عمله.

٤- الكلمات الدالة من التصنيف العشرى العالمى KWUC: وتمثل محاولة الاقتراب من التوحيد حيث تعتمد على المصطلحات التى استخدمها التصنيف العشرى العالمى والذى يمثل أداة واسعة الانتشار فى المكتبات العالمية. وقد فتح ذلك الطريق واسعاً أمام استخدام المصطلحات المقيدة.

التوثيق بالمصطلحات المقيدة :

يقصد بالمصطلحات المقيدة تلك المصطلحات التى اتفق عليها سلفاً وسجلت فى أداة سابقة الإعداد والتجهيز ويعتمد عليها كل من الموثق والمستفيد: الموثق فى اختيار المصطلحات أو المداخل

التى يحلل بها النصوص، والمستفيد فى استرجاع الوثائق. والمصطلحات المقيدة هذه غالباً ما تسجل فى نوعين من الأدوات هما:

- قوائم رؤوس الموضوعات Subject Headings Lists.
- المكانز Thesauri.

وقد ارتبطت قوائم رؤوس الموضوعات بالتوثيق اليدوى، بينما ارتبطت المكانز بالتوثيق الآلى. وقد تطورت المكانز بمعزل عن قوائم رؤوس الموضوعات ونحتت لنفسها مصطلحات بعيدة عن مصطلحاتها وإن أدت نفس وظائفها تقريباً.

وقد بدأت قوائم رؤوس الموضوعات فى الظهور تباعاً منذ أواخر القرن الفارط فى الولايات المتحدة وبعد ذلك فى أنحاء متفرقة من العالم فى القرن العشرين وخاصة فى النصف الثانى منه. ولعله من نافلة القول أن نذكر أن قوائم رؤوس الموضوعات تنقسم إلى قوائم عامة تحاول الإحاطة بجميع جزئيات المعرفة البشرية، وقوائم متخصصة تختص بمجال واحد فقط. وهذه الأخيرة قد تكون قوائم بلدان تدور حول بلد معين.

وقد عرفت المكتبة العربية قوائم رؤوس الموضوعات بفئتيها العامة والمتخصصة منذ نحو ثلاثين عاماً، إلا أن القوائم المتخصصة كانت أسبق ظهوراً من العامة. وعلى الساحة الآن فى المكتبات العربية عدد من تلك القوائم نأتى على أوسعها انتشاراً:

* قائمة رؤوس الموضوعات العربية الكبرى / شعبان عبد العزيز خليفة ومحمد عوض العايدى (ثلاث طبعات مختلفة لثلاثة ناشرين وتدور الصفحات حول ٢٠٠٠ صفحة وعدد الرؤوس حول خمسين ألف رأس).
* رؤوس الموضوعات العربية / إبراهيم الخازندار. ثلاث طبعات فى نحو ٥٠٠ صفحة وعدد الرؤوس يدور حول سبعة آلاف رأس.
* قائمة رؤوس الموضوعات العربية / عمادة شئون المكتبات بجامعة الملك سعود. طبعة واحدة فى ٤٠٠ صفحة ثلاثة آلاف رأس موضوع (تجمدت).
* السعودية : قائمة رؤوس موضوعات للمكتبات ومراكز المعلومات. - الرياض: دار المريخ، ١٩٨١.
* قائمة رؤوس موضوعات علوم الدين الإسلامى - القاهرة: العربى للنشر والتوزيع، ١٩٨٥.

ورغم أن مفهوم المكانز هو مفهوم قديم، إذ هى كلمة يونانية الأصل تعنى المخزن أو الخزانة، فقد ذكر قاموس أكسفورد أن هذا المصطلح بدأ استخدامه عام ١٧٣٦ م كمستودع للمعرفة أو خزانة للمعلومات يشبه فى كثير من الوجوه دوائر المعارف أو الموسوعات. وجاء من بين تعريفاته فى قاموس وبستر أنه «**كتاب يضم مفردات أو معلومات عن حقل معين أو مجموعة مفاهيم، وهو على الأخص معجم مترادفات**».

وقواميس المعانى بمفهوم المكانز فى اللغة العربية قديمة. ولقد بدأ هذه القواميس عثمان بن جنى (ق ٤ هـ: ١٠م) وتبعه ابن سيدة (على بن أحمد) فى المخصص وأبو عبيد القاسم بن سلام فى «الغريب المصنف» وغيرهم.

وفى الغرب كان أول مكنز لغوى فى حدود علمنا هو مكنز بيتر مارك روجيه الذى نشره سنة ١٨٥٢ بعنوان «مكنز الكلمات والجمل الإنجليزية» ويتألف من قسمين أحدهما بالمعانى والآخر هجائى بالمفردات.

أما استخدام المكانز فى مجال المكتبات والمعلومات بالمعنى المتخصص فهو وليد النصف الثانى من القرن العشرين، وربما جاءت أول إشارة إلى ذلك سنة ١٩٥٦، وأول من استخدمت المصطلح فى سياق استرجاع المعلومات هى هيلين برونسون فى مايو سنة ١٩٥٧ وتبعها أو ربما جاء معها فى نفس العام بيترلون الذى طالب بإعداد قاموس معانى يجمع تحت المفهوم الواحد كل المصطلحات الدالة عليه فى «شجرات مفاهيم» يشبه ذلك الذى أعده بيتر مارك روجيه.

ولعل أول مكنز شامل استخدم فى أغراض التوثيق الآلى واسترجاع المعلومات هو ذلك الذى توافرت عليه وزارة الدفاع الأمريكية سنة ١٩٦٠. وقد استند إلى قائمة رؤوس الموضوعات التى كانت تعدها الوزارة من قبل. وقد تلت هذا المكنز مكانز أخرى سنة ١٩٦١م أكثر تخصصاً واستمرت المسيرة حتى يومنا هذا.

ورغم أن قوائم رؤوس الموضوعات العربية قد تخلفت عن نظيراتها فى الغرب نحو ثلاثة أرباع القرن إلا أن المكانز فى اللغة العربية كانت أسعد حظاً، فلم تتخلف أكثر من ربع قرن. ورغم وجود ترجمات عربية مختلفة لكلمة Thesaurus منذ منتصف الستينيات إلا أن الترجمة الحالية «مكنز» أخذت فى الظهور فى الكتابات العربية مع منتصف السبعينيات ويعزى الفضل فى نحت هذا المصطلح العربى الأصيل إلى الزملاء «المغاربيين» وإن لم نستدل على الشخص وجنسيته الذى كان أول من نحته. أما إنتاج مكانز عربية فقد تأخر عقداً على الأقل بعد نحت المصطلح.

ولقد سارت المكانز فى نفس الطريق الذى سارت فيه قوائم رؤوس الموضوعات فثمة مكانز عامة ومكانز متخصصة. وقد يكون المكنز فى هذا الصدد أحادى اللغة أو متعدد اللغات.

أما عن ترتيب المصطلحات داخل المكنز فهناك خط التنظيم الأول وخطوط التنظيم الجانبية، والأصل أن يكون خط التنظيم الأول مصنفاً يعكس العلاقات الهرمية والطبقية بين المصطلحات؛ ويتبعه فى الأعم الأغلب تنظيم جانبى هجائى يعتبر بمثابة الكشاف الهجائى. وقد يكون خط التنظيم الأول هجائياً مع خطوط تنظيم جانبية أخرى مصنفة وغير مصنفة.

وبعض المكانز يقف فيها خطا التنظيم المصنف والهجائى على قدم المساواة يكمل أحدهما الآخر.

❊ ❊ ❊

المكانز فى اللغة العربية :

تتعامل نظم المعلومات الآن مع اللغة العربية والخط العربى بكفاءة، اختزاناً واسترجاعاً وإن لم يخل الأمر من بعض المشكلات التى سنعرض لها فيما بعد. ونتيجة لذلك أنشئ العديد من قواعد البيانات سواء الببليوجرافية أو النصوص الكاملة. وكان من الطبيعى أن يتم تطوير بعض المكانز التى تخدم فى أغراض التوثيق الآلى والاسترجاع باللغة العربية والخط العربى.

وتعتبر جامعة الدول العربية ومنظماتها النوعية هى الحضانة الأساسية لتفريخ الجانب الأكبر من المكانز العربية، ونحن لا نسعى فى هذا البحث إلى حصر كل المحاولات العربية فى هذا الصدد وإنما نحاول رصد الخطوط العامة للمكانز المنشورة على انفراد ومحركات البحث فى بعض قواعد البيانات النصية.

ويعتبر إصدار المواصفة القياسية العربية رقم ٥٧٨ لسنة ١٩٨٤ الخاصة بإعداد المكانز أحادية اللغة، والمواصفة القياسية العربية رقم ٧٩٥ لسنة ١٩٨٥

الخاصة بالمكانز متعددة اللغات (اللتين بنيتا على المواصفتين العالميتين ٥٩٦٣ و ٥٩٦٤ على التوالى) يعتبر صدورهما علامة فارقة فى إعداد المكانز فى اللغة العربية.

من أشمل المكانز العربية «**الجامعة: مكنز ثلاثى اللغات؛ العربية؛ الإنجليزية، الفرنسية**» الذى أصدره مركز التوثيق والمعلومات فى جامعة الدول العربية، سنة ١٩٨٧. وهذا المكنز ليس تأليفاً خالصاً وإنما هو تعريب لمكنز مكتبة الأمم المتحدة (داج همرشولد) وذلك بعد محاولة الاعتماد على عدد من المكانز الأجنبية. ومن الواضح أن الاختيار وقع على هذا المكنز لتشابه اهتمامات كلا المنظمتين وقد جرت إضافة الموضوعات العربية البحتة وحذفت بعض الموضوعات التى لا تخدم المكتبة العربية، كما حدث تعديل فى هيكل المكنز. ويبلغ عدد المصطلحات المستخدمة فيه نيفاً وعشرة آلاف مصطلح. وينقسم مكنز الجامعة إلى أربعة أقسام رئيسية هى:

أ ـ قائمة هجائية مرتبة حسب الواصفات العربية.

ب ـ قائمة مصنفة باللغات الثلاث.

ج ـ كشاف عربى بالكلمات الدالة خارج السياق.

د ـ كشافات باللغات الثلاث.

ويعرض **القسم الهجائى** كل مصطلح فى هيكله الكامل فتشتمل الواصفات على جميع العلاقات (الطبقية أو الهرمية والندية أو التساوى والتداخل أو الترابط) ويتبع ذلك الملاحظات أو التبصرات ورقم الوجه. أما اللاواصفات فإنها تضم المحال إليه (علاقة الندية) ورقم الوجه.

وقد تم ترتيب الواصفات ترتيباً هجائياً طبقاً لقواعد معينة (الشفرة العربية الموحدة، (**أسمو** ٤٤٩)، على أساس طريقة كلمة بكلمة وليس حرفاً بحرف.

وبقيت أداة التعريف فى بداية المدخل رسماً فقط وحذفت حكماً، وثبتت رسماً وحكماً فى أى موضع آخر من الواصفة. كما اعتبرت حروف الجر المرتبطة بالكلمة جزءاً أساسياً منها.

أما **القسم المصنف** فإنه يحدد الفئات الفرعية ويعرض جميع المصطلحات التى تدخل تحت كل فرع. ولا يتضمن هذا القسم العلاقات بين المصطلحات ولا يتبع المصطلح سوى الملاحظات وإحالة «استخدم».

وقسم الكشافات ورد فيه كشافان:

الأول مرتب باللغة الإنجليزية مع المقابلات بالفرنسية والعربية.

الثانى مرتب باللغة الفرنسية مع المقابلات بالإنجليزية والعربية.

وفيما يتعلق بأسماء الأماكن والمنظمات فقد آثر المكنز استخدام الصيغ التى وردت فى المواصفة العربية ٦٤٢ واستخدمت الصيغة الكاملة فى اللغات الثلاث وفى حالة أسماء المنظمات استخدمت الصيغة المختصرة فى اللغتين الإنجليزية والفرنسية حسب دليل مختصرات الأمم المتحدة وأحيل من الاسم الكامل إلى الاسم المختصر. أما بالنسبة لأسماء المنظمات فى اللغة العربية فقد أدخلت بالصيغة الكاملة.

وتتألف الواصفة من كلمات مفردة أو مركبة وجاء ترتيب الكلمات فى الأعم الأغلب بالصيغة الطبيعية. واستخدمت صيغة الجمع فى حالة الواصفات القابلة للعد وفئات وأقسام الأشياء، واستخدمت صيغة المفرد فيما يتعلق بالعمليات والخواص والأشياء المتفردة.

وفى حالة اتفاق الواصفات لفظاً واختلافها دلالة استخدمت حواشى حدية (مقيدات محصورة بين قوسين).

وترد الواصفات والمختصرات فى الجسم الرئيسى على النحو التالى:

الواصفة العربية

الواصفة الإنجليزية

الواصفة الفرنسية

رقم الوجه

س ل : استخدم ل

ت : تبصرة

م ع : المصطلح الأعم (العريض)

م ض : المصطلح الأضيق

م ت : المصطلح المترابط

اللاواصفة

رقم الوجه

س : استخدم

وفى سنة ١٩٩٥ قامت مؤسسة عبد الحميد شومان فى عمان بالأردن بإعداد **المكنز الشامل** الذى بنى أساساً على مكنز الجامعة ولكنه وسع مجاله توسيعاً كبيراً مما جعله أضخم مكنز عام فى اللغة العربية (خمسون ألف مصطلح).

وفى إطار جامعة الدول العربية قام مركز التنمية الاقتصادية بإصدار «المكنز الشامل للمصطلحات فى مجالات التنمية الاقتصادية والاجتماعية» سنة ١٩٧٩ وربما كان هذا المكنز هو أول محاولة فى اللغة العربية لإعداد مكنز. وقد وقع اختيار مركز التنمية الاقتصادية على مكنز **منظمة التعاون الاقتصادى والتنمية فى باريس** الذى شارك فى إعداده سبع وعشرون منظمة دولية وإقليمية وأشرفت على إصداره باللغتين الإنجليزية والفرنسية منظمة التعاون المذكورة ثم صدر بعد ذلك باللغات الإسبانية والألمانية والبرتغالية والفارسية والأندونيسية. والهيئات الرئيسية التى أسهمت فى إعداد هذا المكنز هى:

١- منظمة الأمم المتحدة للأغذية والزراعة بروما.

٢- المؤسسة الألمانية للدول النامية ببون.

٣- اللجنة الدولية للعلوم الاجتماعية والتوثيق بباريس.

٤- مكتب العمل الدولى بجنيف.

٥- منظمة التعاون الاقتصادى والتنمية بباريس.

وكان من بين المنظمات الإقليمية الاثنتين والعشرين التى ساهمت فى إعداد هذا المكنز «**مركز التنمية الصناعية للدول العربية**» الذى تولى وضع الطبعة العربية من المكنز.

وهذا المكنز ليس ترجمة صريحة ولكنه تعريب للمكنز الأجنبى بقصد: «تطويع وإعداد أداة مناظرة تحكم التحليل الموضوعى للوثائق العربية الصادرة فى مجال التنمية الاقتصادية والاجتماعية بشكل عام مع التركيز على التنمية الصناعية».

ويغطى المكنز تسعة عشر مجالاً من مجالات النشاط الاقتصادى الاجتماعى وهى:

١- التعاون الدولى والعلاقات الدولية.

٢- السياسة الاقتصادية والاجتماعية والتخطيط.

٣- الأحوال الاقتصادية والبحث الاقتصادى والنظم الاق×تصادية.

٤- هياكل المؤسسات.

٥- ثقافة المجتمع.

٦- التعليم والتدريب.

٧- الزراعة.

٨- الصناعة.

٩- التجارة والتبادل التجارى.

١٠- النقل.
١١- العلاقات الدولية والتمويل والعملة.
١٢- الإنتاجية والتسيير.
١٣- العمل .
١٤- علم السكان .
١٥- الأحياء والغذاء والصحة.
١٦- البيئة والموارد الطبيعية .
١٧- العلوم الجوية والعلوم الأرضية.
١٨- العلوم والبحث وطرق البحث.
١٩- الإعلام والتوثيق.

وكانت ترجمة هذا المكنز قد بدأت بالفعل سنة ١٩٧٤ ولكن النشر لأسباب مختلفة لم يتم إلا كما رأينا سنة ١٩٧٩ .

وينطوى هذا المكنز على قسمين: القسم المصنف والقسم الهجائى، وفى القسم المصنف تأتى بيانات المصطلحات كاملة: المصطلح ورقم الكود ثم المقابل بالإنجليزية والفرنسية وفروع المصطلح والملحوظة والعلاقات. وفى القسم الهجائى استخدم رقم الكود للربط مع القسم المصنف.

وقد استخدمت اختصارات مختلفة عن تلك المستخدمة فى مكنز الجامعة ولم يتفقا إلا فى واحد فقط وتسير المختصرات فيه على النحو الآتى:

أعم	BT
أخص	NT
متصل	RT
استخدم	USE
بدلاً من	UF

وهذا المكنز بدائى، جنح إلى استخدام صيغة المفرد أكثر من صيغة الجمع. والمكنز مليء بالأخطاء ووجوه القصور.

وفى سنة ١٩٨٩ قامت منظمة العمل العربية ـ المؤسسة العربية للتشغيل بإصدار مكنز (**العمل: مكنز القوى العاملة والتشغيل؛ ثلاثى اللغات: العربية ـ الفرنسية ـ الإنكليزية**) . وهو الآخر ذو قسمين: قسم هجائى وقسم مصنف. ويسير فى نفس الخطوط العامة التى سار عليها مكنز الجامعة.

وفى أوائل التسعينيات قامت المنظمة العربية للثروة المعدنية بإصدار «**المكنز العربى لعلوم الأرض**». ويغطى هذا المكنز واحداً وعشرين موضوعاً رئيسياً تمثل مختلف الميادين التى تم التطرق إليها فى التقارير والمنشورات والخرائط الجيولوجية والمعدنية العربية. ويضم المكنز نحو خمسة آلاف مصطلح عربى وعالمى لتحليل وتخزين المعلومات الجيولوجية والمعدنية من بينها ألف مصطلح خاص بالموضوعات العربية.

وفى سبيل إعداد هذا المكنز تم الاعتماد على العديد من المكانز والتصانيف والمراجع المتخصصة فى الجيولوجيا والمعادن.

وقد استخدمت هنا أيضًا اختصارات مختلفة عن مكنز الجامعة :

م ش	المصطلح الشامل	TG	BT
م ض	المصطلح الضيق	TS	NT
ن ى	انظر أيضًا	VA	SA
س ل	استعمل لـ	EP	UF
س	استعمل		USE

وقد قسم المكنز إلى قسمين: القسم الهجائى الذى يضم الكلمات المفتاحية الجيولوجية العامة مع بيان العلاقات بينها. ويقدم هنا لكل مصطلح ضيق المصطلحات الشاملة بصفة تسلسلية. القسم الثانى يضم المصطلحات الجغرافية والجيولوجية العربية حيث رتبت التكوينات الجيولوجية العربية هجائياً مع بيان العلاقات بينها وبين الأعمار الجيولوجية العامة التى تمثل المصطلحات الشاملة لها مع إعطاء علاقة الترابط مع البلد أو البلدان التى توجد فيها التكوينات.

والموضوعات التى يغطيها المكنز تسير على النحو التالى:

٠١ـ عموميات
٠٢ـ طرق (أو مناهج)
٠٣ـ تحليل رياضى
٠٤ـ الخواص الكيميائية والفيزيائية
٠٥ـ الجيوكيمياء
٠٦ـ علم المكامن
٠٧ـ الاقتصاد المنجمى
٠٨ـ المواد الصناعية
٠٩ـ علم المعادن
١٠ـ علم الصخور (صخور متبلورة ـ علم الصخور الرسوبية)
١١ـ أوسيانوغرافيا
١٢ـ أستراتيغرافيا
١٣ـ الجيولوجيا البنيوية
١٤ـ فيزياء الأرض
١٥ـ الجيوفيزياء
١٦ـ هيدرولوجيا
١٧ـ هدروجيولوجيا
١٨ـ جيولوجيا هندسية
١٩ـ جيومورفولوجيا
٢٠ـ علم التربة
٢١ـ علم الأحياء القديمة

وخارج إطار جامعة الدول العربية تم إعداد عدد من المكانز من بينها ما هو منشور ومن بينها ما هو حبيس قواعد البيانات التى أعد من أجلها. ومن تلك المنشورة (**مكنز الفيصل: مكنز عربى شامل فى علوم الحضارة: قسم علوم الدين الإسلامى**) والذى نشره مركز الملك فيصل للبحوث والدراسات الإسلامية سنة ١٩٩٤، وينقسم هو الآخر إلى قسمين قسم للعرض الهرمى (المصنف) وقسم للعرض التباديلى (الهجائى).

والمجالات الأساسية فيه تسير على النحو التالى:

ق / ب	علوم الدين الإسلامى
ب	مصطلحات عامة
ج	علوم القرآن
د	علوم الحديث
س	سيرة
ص	أصول الفقه
ع / ط	فقه
ف	تصوف
ق	عقائد

وعلى سبيل المثال تسير المصطلحات العامة على النحو الآتى:

مبادئ الإسلام	٥ ب
الدعوة الإسلامية	١٠ ب
ثقافة إسلامية	١٥ ب
تربية إسلامية	٢٠ ب
مساجد	٢٥ ب

أوقاف	ب	٣٠
الإسلام والعلم	ب	٣٥
استشراق	ب	٤٠
مذاهب فكرية	ب	٥٠
مقارنة الأديان	ب	٦٠
فلسفة إسلامية	ب	٦٥
فلاسفة إسلاميون	ب	٧٠
قضايا معاصرة	ب	٧٥
الأوقات	ب	٨٠

وحيث عالجها الزميل مساعد بن صالح الطيار بشيء من التفصيل فى بحثه (انظر المصادر) فسوف نلم بها هنا فقط إلماماً سريعاً.

تعتبر الموسوعة الذهبية أكبر مستودع للحديث النبوى الشريف حيث حملت عليها ١٢٨ كتاباً من أمهات كتب الحديث النبوى وشروحها وبلغ عدد المجلدات ٤٠٠ مجلد، وعدد الروايات فيها ٢٠٠ ألف رواية و٨٠ ألف حكم على الأحاديث. والكشافات فيها متنوعة حيث يمكن الوصول إلى محتوياتها عن طريق راوى الحديث والكلمات المفردة والجذور وعدة كلمات. ويمكن لمستخدم النظام التعليق عليه، كما يمكن القص واللصق والطبع.

وعلى جانب المكانز حبيسة قواعد البيانات، أى التى لم تنشر مطبوعة فقد صاحبت هذه المكانز حركة النشر الإلكترونى، وخاصة لما يعرف بالكتب المرجعية، وهناك العشرات من تلك الأعمال المرجعية ذات النصوص الكاملة التى يحتاج البحث فيها إلى مكانز أو أدوات بحث ذاتية مبنية مع النصوص نفسها، ومن بين تلك الأعمال نذكر:

١- الموسوعة الذهبية	التراث	١٤١٨ هـ	١٩٩٧ م
٢- موسوعة الكتب التسعة	العالمية	١٤١٦ هـ	١٩٩٥ م
٣- برنامج السنن الأربع	السيليكون	١٤١٦ هـ	١٩٩٥ م
٤- موسوعة الفقه الإسلامى	العريس	١٤١٧ هـ	١٩٩٦ م
٥- فقه المعاملات	العالمية	١٤١٦ هـ	١٩٩٥ م
٦- برنامج المغنى	البرامج الإسلامية	١٤١٥ هـ	١٩٩٤ م
٧- برنامج حياة الصحابة	المعالم	١٤١٥ هـ	١٩٩٧ م
٨- مكتبة البيت المسلم	الخطيب	١٤١٨ هـ	١٩٩٧ م
٩- لسان العرب	المستقبل	١٤١٦ هـ	١٩٩٥ م
١٠- موسوعة المورد الإلكترونى	المستقبل	١٤١٦ هـ	١٩٩٥ م

وموسوعة الكتب التسعة هى الأخرى كسابقتها حملت عليها أحاديث الكتب التسعة الصحاح: صحيح البخارى، صحيح مسلم، سنن أبى داود، سنن الترمذى، سنن النسائى، سنن ابن ماجه، موطأ مالك، سنن أحمد، سنن الدارمى. وقد بلغ عدد الأحاديث هنا نحو ثلاثة وستين ألف حديث، وهذه الموسوعة متنوعة الكشافات حيث يوجد نحو ٤٥٠٠ رأس موضوع إضافة إلى أسماء الأماكن والأشخاص.

وبرنامج السنن الأربع هو الآخر يتضمن أحاديث

كتب السنن الأربعة إضافة إلي أحكام الشيخ الألبانى عليها من حيث درجة الثقة فى كل حديث، ويرى بعض الثقاة أن هذه الإضافة قد أضفت قيمة عالية عليها. ويمكن البحث فى هذه القاعدة عن طريق الكلمات المفردة وعدة كلمات مركبة، كما يمكن البحث عن طريق جذر الكلمة وموضوع الحديث، وكذلك يكفل النظام استعراض قائمة محتويات كل كتاب على حدة كما وردت فى الأصل المطبوع. كما يؤدى النظام خدمات الطباعة وحفظ نتائج البحث.

ومما لاشك فيه أنه يمكننا القول مطمئنين أن الأحاديث النبوية قد تمت تغطيتها بتوثيق آلى واسع النطاق.

وقد حظى الفقه الإسلامى هو الآخر بنصيب لا بأس به من التوثيق الآلى كذلك، فهناك مكتبة الفقه الإسلامى التى تضم أكثر من مائتى مجلد فى الفقه على المذاهب الأربعة الأشهر (الحنفى ـ المالكى ـ الشافعى ـ الحنبلى). ويتم البحث فيها عن طريق التجذير وكشافات الآيات القرآنية والأحاديث والأسماء وغيرها. وينطوى البرنامج على إمكانيات تخزين وطباعة النصوص.

أما برنامج **فقه المعاملات** فهو يدور حول جزء من الفقه وهو عبارة عن فصول مختارة وأبواب من كتب الفقه العامة ولكنها متخصصة فقط فى فقه المعاملات. ولما كان ناشر هذه القاعدة هو نفسه ناشر موسوعة الكتب التسعة فى الحديث فقط استخدم نفس محركات البحث؛ وهى البحث بالكلمة والجذور ورؤوس الموضوعات وأسماء الأماكن والأشخاص والمدارس الفقهية والآيات القرآنية والأحاديث.

وبرنامج **كتاب المغنى** يدخل هو الآخر فى باب الأعمال الفقهية إذ هو موسوعة علمية ضخمة فى الفقة المقارن والبحث فيه عن طريق عناوين المسائل الفقهية. وعموماً فإن محركات البحث فيه ضعيفة مهزوزة.

وبرنامج **حياة الصحابة** يدخل فى باب قواعد المعلومات الببيوجرافية أو السير ويتضمن معلومات عن صحابة رسول الله صلى الله عليه وسلم. ومحركات البحث الأساسية هنا الروابط المنطقية مع إمكانية البحث من خلال النص وقائمة محتويات الكتاب بأجزائه الثلاثة.

وأما **مكتبة البيت المسلم** فهى سلسلة من الكتب تضم فى حلقاتها الأولى سبعة وستين كتاباً فى علوم القرآن والحديث والسيرة النبوية واللغة العربية والنحو، وبعض الكتب تقع فى عدد من المجلدات. ومحركات البحث الأساسية هنا هى كلمات النص وقوائم محتويات الكتب. ويتيح البرنامج إمكانية طباعة النصوص.

ولسان العرب هو المعجم اللغوى الأشهر لابن منظور. ويتم البحث فيه كما يتم فى النسخة المطبوعة على الحروف الهجائية، كما يوفر طريقة البحث عن طريق الجذور وكذلك يوفر البرنامج نظاماً لنطق الحروف والجذور لمعرفة النطق الصحيح.

وموسوعة المورد هى دائرة معارف عربية مطبوعة فى اثنى عشر مجلداً حملت فى قاعدة واحدة ويدور عدد مداخلها حول ١٥٠٠٠ مادة فى فروع المعرفة البشرية ولكن المواد عموماً قصيرة أقرب إلى المعاجم الموسوعية. وهذه الموسوعة تؤمن النص والصوت والصورة فى آن واحد. وليس هناك طريقة البحث الحر فيها وإنما يتم الدخول إلى كل مقالة عن طريق عنوانها حيث يتم استعراضه. ويمكن الوصول إلى المقالات كذلك عن طريق الموضوعات العريضة: علماء فلاسفة، تاريخ، جغرافيا، طب.

ورغم أن الخط العربى واللغة العربية يستخدمان الآن فى قواعد البيانات وقواعد النصوص الكاملة اختزانًا واسترجاعًا ومحركات البحث، إلا أنه مانزال هناك بعض مشكلات التعامل التى من بينها مشكلات فنية بحتة ومشكلات لغوية بحتة ويأتى على رأسها:

١- عدد الحروف العربية ٢٨ حرفًا بينها حرف واحد متحرك (الألف) وحرفان شبه متحركين (الواو والياء)، حيث فى بعض المواضع تستخدمان كحرف متحرك وفى مواضع أخرى كحرف أصلى. يضاف إلى ذلك علامات التشكيل الخمس التى تستخدم كعلامات إعراب أحيانًا مما يزيد من قائمة الرموز اللازمة فى الكتابة العربية.

٢- تبدأ الكتابة العربية من اليمين إلى اليسار عكس الخط اللاتينى وهذا الاتجاه فى الكتابة يمثل سلسلة من المشكلات فى قواعد البيانات وخاصة متعددة الخطوط.

٣- تغير شكل الحرف الواحد طبقًا لوضع الحرف داخل الكلمة: أولها، داخلها، آخرها. وبعض الحروف كما نرى له ثلاثة وأحيانًا أربعة أشكال. وربما كانت الهمزة من أعقد الحروف العربية فى هذا الصدد فهى تعتبر حينًا حرفًا صامتًا وأحيانًا علامة نطق. والهمزة قد تأتى فوق الألف أو تحته، كما قد تأتى على السطر وقد تأتى على الواو وعلى الياء وقد تكون ممدودة لتصبح ألفين وفى هذا الصدد فإن القيم الصوتية فى الخط العربى تصل إلى ٤٣ قيمة مقارنة بها فى الخط اللاتينى الذى يمثل ٢٦ فقط وحيث لا تتغير قيمة الحرف الواحد هناك.

٤- هناك مشكلة أخرى تتعلق بالأرقام حيث تكتب بشكلين أرقام الغبار أو التخت 1, 2, 3, 4, 5. 6.7.8.9 والشكل الهندى١، ٢، ٣، ٤، ٥، ٦، ٧، ٨، ٩، ٠ وأيًا كانت أشكال الأرقام فإنها لابد وأن تكتب من اليمين إلى اليسار وإن كانت حاليًا تقرأ من اليسار إلى اليمين فى حالة الرقم.

٥- الافتقار إلى الحروف المتحركة فى جذور الكلمات العربية وخاصة الأفعال وغالبيتها ثلاثية الجذر. وربما يختلط الفعل مع الاسم مثل ك ت ب ومع إضافة بعض الحروف قد يختلف المعنى: مكتب، مكتبة. وربما لاتكون هناك مشكلة فى هذا مع البشر ولكنها تمثل مشكلة حقيقية مع الحاسبات الآلية.

٦- الضمائر التى تلحق بأواخر الكلمات والتى تغير أواخر الكلمات كتابى، كتابها، كتابهم، كتابهما، كتابهن...

٧- هناك أيضًا مشكلة ضبط الإملاء حيث إن بعض حروف العطف والجر قد تلتصق بالكلمة مثل واو العطف وباء الجر وكاف الجر... حيث يجب أن تكون هناك مسافة بين الحرف والكلمة الأصلية حتى تنطق الكلمة بعيدة عن الحرف.

❊ ❊ ❊

مستقبل التوثيق الآلى:

يقول جيرالد سالتون إن استرجاع المعلومات من قواعد المعلومات الآلية عن طريق اللغة الطبيعية بدون قيود وبلا حدود هو أمر مستبعد على الأقل فى المستقبل المنظور، خاصة وأنه ليس هناك اتفاق على الطريقة المثلى لقولبة محتويات الوثائق وتحليلها بل وحول المعرفة البشرية نفسها خارج المعرفة المتخصصة فى موضوع ما على الأقل. ذلك الاتفاق المطلوب لفهم النصوص وفهم جمل اللغة الطبيعية ووضع استراتيجيات البحث واسترجاع المعلومات.

وقد خرجت دراسات أخرى بنفس النتيجة

تقريباً «إن التكشيف الآلى الكامل مسألة غير عملية وغير ممكنة فى قواعد البيانات الكبيرة فى المستقبل المنظور». وقد أثبتت التجارب فى العقود الأربعة المنصرمة أن الإعداد الآلى لكشافات النصوص فى الكتب والأبحاث والتقارير وغيرها من النصوص الطويلة لم يفلح ولم ينجح نجاح الإعداد اليدوى بدليل أن قواعد البيانات الكبرى التى تغطى مجالات واسعة مثل مستخلصات الكيمياء CA والمستخلصات الحيوية BIOSIS والمستخلصات الطبية MEDLINE وغيرها لا تستخدم أى نظام آلى كامل فى تكشيف ملايين التسجيلات الببليوجرافية. وبدلاً من ذلك تعتمد على فرق عمل مدربة من المكشفين البشر، يدعم عملهم الحاسب الآلى، الذى يساعد فقط فى توفير الوقت لتنفيذ الأعمال التى تتطلب التكرار والجهد العضلى والبحث عن الكلمات فى المكانز فقط وأسماء المركبات الكيميائية وأسماء أعضاء الجسم البشرى وأسماء الأماكن وغير ذلك عن طريق الأمور التى تتطلب التوحيد واستخدام قوائم الاستناد، بل إن بعض نظم التوثيق مثل نظام المكتبة القومية الزراعية الأمريكية قد وجد أن استخدام الحاسب فى مسح النصوص واستخراج المصطلحات منها من المكانز يستغرق وقتاً أطول من البحث البشرى؛ بل وأكثر من هذا فإن المصطلحات المستخرجة من مكانز الحاسب الآلى تحتاج إلى عمليات تحرير من جانب المكشفين أنفسهم.

إن المشكلة الرئيسية في التكشيف والترجمة الآلية تكمن فى أن الآلات الإلكترونية أو الميكانيكية يمكنها فقط أن تطابق أو تضاهى الكلمات مع النص ويمكنها أن تسترجعها بطريقة أو بأخرى ولكنها لا تستطيع أن تفهمها: إنها فقط، تقرؤها. إن ملكة البشر فى الفهم واستيعاب المعانى والسياق، إنما هى عملية تتأتى من معرفة العالم، إنها خليط من التجارب والذكريات والعلاقات والغرائز لا يملكها أى حاسب إلكترونى حتى الآن.

إن علينا أن ننتظر ما قد يأتى به القرن الحادى والعشرون فى مجال التوثيق الآلى، وقد يطول هذا الانتظار.

* * *

بعض المصادر

إضافة إلى المكانز نفسها التى جرى فحصها ثمة بعض المصادر التى يمكن الرجوع إليها للاستزادة ومن بينها:

١ ـ أمنية مصطفى صادق. الأسس والملامح الرئيسية لمكانز العلوم الاجتماعية. مجلة المكتبات والمعلومات العربية. المجلد الخامس، العدد الرابع، أكتوبر ١٩٨٥.

٢ ـ جامعة الدول العربية. الأمانة العامة. مركز التوثيق والمعلومات. بناء المكانز وتطويرها. تونس: الجامعة، ١٩٨٧.

٣ ـ حسين الهبايلى. المكانز متعددة اللغات ومنهجية إعدادها. المجلة العربية للمعلومات، مج٨، ع١، ١٩٨٧.

٤ ـ شوقى سالم. المكانز المتخصصة: نظرة على بعض جوانب إعدادها. المجلة العربية للمعلومات. مج٢، ع٤ يونية، ١٩٨٠.

٥ ـ مساعد بن صالح الطيار. الكتب الإلكترونية العربية: دراسة استطلاعية. دراسات عربية فى المكتبات وعلم المعلومات. مج٤، ع١ يناير ١٩٩٩.

٦ ـ محمد تيسير درويش. المكنز ودورة كركيزة

فنية أساسية لاستخدام الحاسب الإلكترونى «الكمبيوتر» فى خزن واسترجاع المعلومات. رسالة المكتبة ـ مج ١٩، ع٣، سبتمبر ١٩٨٤.

7 - Attchison, Jean. Thesaurus construction; apractical manual..- London, Aslib, 1972.

8 - Borko, Harold and Charles Bernier. Indexing concepts and methods. New York: Academic press, 1978.

9 - Foskett, Douglas J. Thesaurus. in Encylopedia of Library and Information Science. New York: Marcel Dekker, 1980. Vol. 30.

10 - Gilchrist, Alan. The role of thesauri in mechanized systems. Indexer, Vol. 9. Oct. 1979.

11 - Gilchrist, Alan. The thesaurus in retrieval. London: Aslib, 1971.

12 - Ghose, A. Problems of thesaurus construction. JASIS, Vol. 28, No. 4 1977.

13 - Wall, E. Information retrieval thesauri. New York: Engineers Joint Council, 1962.

14 - Willisch, Hans H. Indexing from A to Z. 2 nd ed. - New York: Wilson, 1995.

❄ ❄ ❄