الفصل السادس

قواعد بيانات النص وإدارة الوثائق

المقـــدمة

لقد لقيت نظم إدارة قواعد البيانات الهيكلية وخاصة العلاقية RDBMS كثيرا من الإهتمام خلال السنوات الأخيرة الماضية ، وخاصة عند استعراض البيانات الهيكلية المتمثلة فى صفوف وأعمدة تتسلسل فيها الحروف والبيانات الرقمية والتواريخ فى هذه القواعد العلاقية. إلا أن ذلك يمثل جزءً صغيرًا فقط من المعلومات التى تتوافر فى أى منظمة أو مؤسسة ، ومن المحتمل أن يكون ذلك فى حدود أقل من ١٠٪ من حجم البيانات الإجمالى فى المنظمة.

وفى الحقيقة ، نجد أن النص الموجود فى شكل الوثائق المختلفة من أى نوع ، يعتبر من مصادر المعلومات الأكثر شيوعًا واستخدامًا فى الوقت الحالى ، حيث يتواجد فى شكل المراسلات ، محاضر الجلسات ، التقارير ، رسائل البريد الإلكترونية ، أدلة الإجراءات ، مواد التسويق ، الكتب ، مقالات الجرائد والدوريات . . . إلخ التى نتعامل معها كل يوم . وتوجد ثروة من المعلومات فى هذه الوثائق التى تعتبر كل وحدة بيانات منها مفيدة مثل البيانات الهيكلية فى قواعد البيانات العلاقية وغيرها التى يتوصل إليها مستخدمو الأعمال فى المنظمات المختلفة على أساس يومى .

وتحتاج معظم أو كل منظمات الأعمال والمكتبات ومراكز المعلومات إلى تخزين سجلات الوثائق والنصوص فى قاعدة بيانات النص Text database أو قاعدة النص Textbase طبقًا لكمية الوثائق المنتجة فيها . وفى العادة تكتب الوثائق بطرق متشابهة مع رؤوس Headers وأقسام فرعية Subdivisions ، حيث تتضمن الرؤوس الكلمات الرئيسية Keywords والأكواد Codes التى تسمح البحث للإشارة المضافة إلى الوثيقة المعينة .

كما يمكن أن تتواجد الكلمات الرئيسية فى الوثائق نفسها . وفى هذه الحالة يجب أن يكشف Index نص الوثيقة مع ترك الألفاظ غير المهمة وغير الضرورية . وفى هذا الإطار قد تستخدم أساليب البحث معتمدة على استخدام الألفاظ البولينية Boolean terms لوصل الكلمات والعبارات التى تم تكشيفها من قبل معًا لتوضيح البحث .

كما قد يعتبر تكامل الوصول إلى النص والبيانات الهيكلية من الطرق المساعدة فى تعزيز مدى بيانات المؤسسة الداخلية بطريقة جيدة لترشيد عملية اتخاذ القرارات وتقديم المزايا التنافسية المحتاج إليها .

طبيعة بيانات النص

تشتمل الوثائق أو المعلومات النصية على بيانات ترتبط بتسلسل الحروف المعدة بأسلوب شفرة أسكى ASCII ، ولكنها أكثر تعقيدًا مما هى عليه فى الواقع . علما بأن الحروف فى الوثائق تتكامل فى كلمات ، عبارات ، وجمل . ويتضمن النص كله سلسلة من الخصائص الأخرى التى تشتمل على الشكل (البنط والحجم) ، والهيكلة (الفقرات والأنماط الأخرى مثل العنوان ، رأس الموضوع . . . إلخ) ، والمعنى أو المحتوى . ومن الواضح أن المستخدمين يحتاجون طرقًا للبحث فى محتوى النص تتعدى وظائف تسلسل الحروف التى تقدم بواسطة منتجات نظم إدارة قواعد البيانات .

ويلاحظ أن منتجات البحث فى الوثيقة لا تحل مشكلة الوصول إلى كم كبير من المعلومات ، حيث أنه فى الماضى كان يلبى هذا الإحتياج بواسطة منتجات بحث النص الكامل التى منها أدوات بحث مثل :

- Fulcrum's Search Server
- Open Text's Text Search
- Verity's Topic
- . . . . إلخ .

التى صممتها بيوت إنتاج قواعد البيانات لتكشيف محتويات مجموعة من الوثائق . وتسمح هذه المنتجات للمستخدم أن ينشئ استفسارات متنوعة عن النص ، ويسترجع الوثائق بكفاءة تضاهى معايير البحث الأكثر إرتباطًا . وكلما إزداد نضج هذه المنتجات أو الأدوات كلما إزدادت مستويات التعقيد فى أساليب التكشيف والاسترجاع بالإضافة إلى تحسينات الأداء .

وفى هذا الصدد ، يمكن تتبع تطورين رئيسيين يحتمان على المستخدم رؤية النص كجزء من تكامل أكبر ، وهذان التطوران هما :

١ - حجم النص الكامل وإمتداده ، وعلى وجه الخصوص النص المتوافر بالفعل فى شكل إلكترونى كما هو متواجد على شبكة الويب عند التوسع فيه . وكما ينمو حجم

النص المطلوب البحث فيه تصبح قابلية الإستطراد أو التسلق Scalability والأداء Perfomance من القضايا المهمة .

٢ - إحتياج المستخدمين إلى القدرة فى إعادة استخدام معلومات النص وعرضها مع أشكال أخرى من البيانات . وأحد الأمثلة المرتبطة بذلك فهارس المكتبات ، وفهارس طلب المشتريات البريدية ، حيث يرتبط النص بأشكال كل وثيقة أو كل منتج ، ومع توفير بيانات هيكلية مثل رقم الوثيقة ، عنوانها ، ناشرها . . . إلخ ، أو رقم المنتج وسعره. وبذلك يعتبر النص مجموعة من البيانات الواقعية التى تحتاج إلى إدارة قاعدة بيانات بدرجة كبيرة من الأمن والرقابة المتزامنة والاستناد والاسترجاع وغير ذلك من الخصائص المرتبطة بنظم إدارة قواعد البيانات .

تعريف قاعدة النص

الجمع بين نظام إدارة قاعدة البيانات DBMS مع خصائص ووظائف إسترجاع النص ينشئ ما يمكن أن يطلق عليه قاعدة النص . وترتبط قاعدة النص بما وراء بحث النص Text Search التى تشتمل على القدرة فى تعريف وحدات بيانات هيكلية قليلة فقط ، حيث يكون توجه إدارة بيانات النص فى نمط متوافق وموثوق منه متكامل مع البيانات العلاقية الهيكلية فى وسيلة استرجاع فردية مما يعتبر مفتاح الإختلاف بين بحث النص الكامل ونظام قاعدة النص . من هذا المنطلق ، يمكن تحديد السيناريو الذى يفسر كيف تعمل قاعدة النص . وعلى سبيل المثال ، فى نطاق حزمة برمجيات «ميكروسوفت ورد MS. Word» التى تستخدم مع برامج النوافذ Windows ، يختار المستخدم خيار «فتح Open» من على قائمة الملف File Menu ويوضح الوثيقة المسجلة فى دليل محلى . على أنه فى إطار معالج النص بدل القراءة من مشغل الطلب المحلى ، فإنه يقوم بإصدار عبارات إختيار فى مواجهة جدول متواجد على شبكة كمبيوتر محلية LAN لاسترجاع الوثيقة بشفافية ، حيث تكون الوثيقة فى الواقع عموداً فى صف ، مجموعة صفوف فى قاعدة البيانات ، أو ملف يمكن الوصول إليه خلال مؤشر فى قاعدة البيانات ، وبذلك تصبح قاعدة بيانات النص مستودع معلومات كامل للمؤسسة . وفى إمكان أى مستخدم إصدار التساؤلات التى يمكنها الوصول إلى البيانات فى قاعدة البيانات . ويوضح ذلك الإختلاف الرئيسى بين معالجة البيانات محليًا والتعاون والمشاركة فى شبكة كمبيوتر ، حيث أن من الأشياء التى تنجزها قاعدة البيانات جيداً ، ما يرتبط بالمساعدة فى مشاركة المعلومات وإدارتها ويساهم فى أداء هذه العمليات مجموعة من الخصائص المرتبطة بالأمن ، ضبط التزامن ، وإدارة أى تصرف معين .

وفى الخطوة التالية يمتد هذا الحل لأنواع البيانات الأخرى ، حيث تطبق أيضًا متطلبات العميل أو المستخدم على أنواع البيانات الثنائية Binary الأخرى المعقدة كما فى الأشكال والبيانات المتزامنة Synchronous كبيانات التدفق المرتبطة بأطر الفيديو أو الأطر السمعية التى يجب أن تتناسق وتمتد فى نطاق فترة زمنية معينة . وفى هذا السياق يفترض حاجة المستخدمين لمجموعة من الأدوات والخدمات التى تعمل على تكامل الوصول إلى النص والبيانات الهيكلية مع أنواع البيانات الأخرى . ويتطلب ذلك ما يلى :

١ - الحاجة لتخزين كل أنواع البيانات فى قاعدة بيانات منطقية تشتمل على بيانات هيكلية ، وبيانات النص والأشكال والفيديو والصوت ، مع تتابعات الوقت وأنواع البيانات الأخرى المفسرة للمستخدم ، مما يسمح بتقدير علاقات أساسية بطريقة أسهل بين البيانات .

٢ - لغة مفردة مثل لغة الاستفسار أو التساؤل الهيكلية SQL الممتدة والتى يمكنها الوصول إلى كل أنواع البيانات . ويرتبط الغرض من ذلك فى التوصل إلى القدرة فى إصدار تساؤل واحد يمكنه التعبير عن معايير البحث وتجميعها عبر كل أنواع البيانات كما يعيد مجموعة واحدة من النتائج المتكاملة ، مما يتطلب كثير من الحلول لبحث النص وخاصة عندما يصدر المستخدم إستفسارين أو تساؤلين : أحدهما لإسترجاع النص المعين المطلوب من مصدر النص ، والآخر لاسترجاع البيانات الهيكلية وأنواع البيانات الأخرى كالأشكال عندما تخزن فى قاعدة بيانات منفصلة . عندئذ تعتبر حزمة برمجيات الحاسب العميل Client مسئولة عن تجميع النتائج .

٣ - إشتمال الحلول الكاملة على وظائف الحاسب الخادم Server ، وسهولة استخدام أدوات العمل للاستفسار ، والتقرير ، وتطوير التطبيق ، وأدوات الإدارة .

وفى هذا الإطار يوجد مصطلحان أساسيان هما : قاعدة النص Text base وقاعدة الوثيقة Document base . وتعرف قاعدة النص بأنها تتضمن النص فقط الذى تخزن فيه وحدات البيانات وتمثل كتلا من النص معروفة باسم معين . ويمكن أن يكون هذا الاسم بسيط كرقم تزويد Accession No. ، أو قد يحتوى على معنى معين . ولا يتطلب أن تشتمل الوثائق على الهيكل نفسه .

أما مصطلح قاعدة الوثيقة فيمثل النص الذى يرد مرتبا فى شكل هيكلى . وفى العادة تتضمن الوثائق سجل رئيسى Header Record الذى يعطى بعض المعلومات عن الوثيقة والنص الفعلى بها . ومعظم الوثائق تكون مرتبة هرميًا من هيكل العنوان . وتتفرع الكتب إلى فصول ، والفصول إلى فقرات والفقرات إلى جمل ، وأخيراً تتفرع الجمل إلى كلمات وهكذا . وقد تتضمن الوثائق رسومات وتوضيحات مختلفة . ويمكن البحث عن الوثيقة إما بالسجل الرئيسى الخاص بها أو البحث عن محتواها . ويعتبر السجل الرئيسى الخاص

بالوثيقة مشتملاً على مجموعة من الحقول الثابتة التى تطبق على كل وثيقة متوافرة فى النظام. ويمكن أن تستخدم قاعدة بيانات علاقية فى البحث عن ذلك بطريقة أسرع من بحث النص ، كما يمكن أن يشتمل السجل الرئيسى للوثيقة على كثير من الحقول التى تصف الوثيقة ماديا وموضوعيا . ومن أكثر الحقول شيوعا ، حقول العنوان ، المؤلف أو المؤلفون ، بيانات النشر والتاريخ ، رقم التصنيف ، الكلمات الرئيسية ، أو مستخلص الوثيقة . . . إلخ. وتستخدم معظم حزم البرمجيات الخاصة بذلك قائمة تسلسل الحروف المنفصلة للحقول ذات القيم المتعددة للمؤلفين وقوائم الكلمات الرئيسية أى المكانز Thesausi التى قد تشتمل على أكواد لبحث التتابعات المختلفة .

قدرات بحث النص المتقدمة

يحتاج النص إلى متطلبات تكنولوجية لتداول بيانات النص . ومن القدرات المتقدمة فى بحث النص ، ما يلى :

١ - بحث النص الكامل : Full - Text Search

تعتبر هذه القدرة محور تكنولوجيا بحث النص ، كما ترتبط بالجهود الرئيسية لمخاطبة متطلبات استرجاع النص المتمثلة فى محركات بحث النص الكامل Full - Text Search Engines . وتعمل هذه الأدوات على إسترجاع الوثائق المبنية على الكلمات والعبارات أو المفاهيم المتضمنة فى الوثائق والتى كانت متاحة منذ فترة طويلة من قبل . إلا أنه قد ظهر إلى الوجود فى السنوات الحديثة أدوات قوية لبحث النص ترتبط بأساليب التكشيف والقدرة على تقرير مدى مطابقة النص مع استفسارات المستخدم . والهدف الرئيسى من ذلك يتمثل فى مساعدة المستخدم العثور على كل الوثائق المرتبطة باستفساراته أو تساؤلاته المختلفة فقط، بدلاً من إمداد بكل الوثائق أو استرجاع الوثائق غير المتصلة بموضوع بحثه أو استفساره.

٢ - أساليب تكشيف الكلمة : Word Indexing Techniques

يحدد المستخدم فى نطاق بحث النص الكامل كلمة أو أكثر من كلمة التى يجب أن تتضمن فى الوثيقة المعينة التى تعرض فى قائمة نشطة Hit List. . وتستخدم معظم «محركات بحث النص الكامل» تكنولوجيا التكشيف نفسها ، منتجة ما يطلق عليه «قائمة الكلمات المعكوسة Inverted Word List» لأغراض التكشيف . ويوجد مفهومان أساسيَّان ، هما :

- **المفهوم الأول** يتمثل فى توفير «قائمة توقف Stop List» التى تعرِّف كل الكلمات غير المكشفة أو غير ممكنة البحث بواسطة المستخدم. وأمثلة كلمات التوقف ، الشولة (،) أدوات التعريف مثل «الـ The» وحروف المفرد مثل «a, an»، وحروف الجر مثل «لـ to» . . . إلخ . ولا تكشف كلمات التوقف كما لا تتضمن فى البحث .

- **المفهوم الثانى** يرتبط بقائمة الأداء الفعلى الذى يشتمل على كل الكلمات الممكن البحث عنها .

ويتمثل أحد مساوئ استخدام كلمات التوقف فى عدم قدرة محرك البحث فى العثور عليها كما فى عبارات مثل «الارتباط مع Association of» أو «العائد من الاستثمار Return on Investmant» . ولذلك يجب أن يقدم منتج البحث الكامل فرصة الإختيار المرتبطة بتحديد كلمات التوقف أو عدم تحديدها .

أما مجموعة الكلمات المكشفة أو الممكن البحث من خلالها ، فإنها تقتصر على مجموعة من الكلمات التى تعتبر فريدة فى حد ذاتها والتى تختصر للجذور أو الكلمات الأصلية المصحوبة بمؤشرات تحدد مواقعها فى الوثائق المرتبطة بها . وعلى ذلك تحتاج كل كلمة أن تخزن مرة واحدة فقط فى الكشاف مع تحديد عدد إحداثياتها Occuronces ومواقعها فى كل وثيقة . فعلى سبيل المثال ، يمكن اختزال وتقليل ما يقرب من مليون كلمة إلى عشرين ألف كلمة فريدة فقط ، على أن يتسم ذلك بالمرونة الكبيرة فى حفظ ومراجعة أو صيانة الكشاف ذاته كلما أضيفت وثائق جديدة أو استبعدت وثائق أخرى من رصيد قاعدة بيانات النص .

وقد تستخدم فى بعض منتجات برمجيات قواعد بيانات النص أساليب تكشيف أخرى مثل أمر «فتح النص Open Text» الذى يشتمل على تسلسل وحدات بتات Bits بدلاً من الكلمات الفردية . ويعتبر ذلك مدخلا يسهل المساندة للكلمات الأجنبية وخاصة لمجموعات الحرف Byte المزدوجة . وفى هذا الإطار ، تستخدم شركة آى.بى.إم IBM مدخلا مبنيًا على الإحصائيات فى التكشيف . كما تستخدم «أداة خادم نص شركة أوراكل Oracle's Text Server» كشافات خرائط النقاط Bitmapped لتحديد مواقع الكلمة . وتقدم تكنولوجيا السياق Context الخاصة بها تحليل لغوى بالإضافة إلى تكشيف الكلمات المعكوس .

٣ - أساليب البحث للعثور على كل النص المتوافق :

تهدف خصائص بحث النص الكامل المهمة الأخرى للعثور على كل الوثائق التى قد

تضاهى معايير البحث التى تتعدى الكلمات الفردية المحددة بواسطة المستخدم . ومن هذه الأساليب :

(١) **المكنز** Thesaurus حيث يستخدم محرك البحث مكنز داخلى لكى يستدعى المترادفات لكلمات أو ألفاظ البحث فى صالح المستخدم . وعلى ذلك ، لا يجب على المستخدم أن يعرف أو يتذكر تحديد كل الكلمات المرتبطة المتفقة فى تساؤل الباحث ، على سبيل المثال ، فى الإستفسار الذى يشتمل على كلمة «سيارة Car» ، قد يبحث «محرك البحث» على كلمات أخرى مثل «أتوموبيل Automobile» ، «مركبة Vehicle» . . . إلخ .

(٢) **الكروت الجامحة** Wild cards ، التى تغطى «الكلمات المبتورة Truncated words ، مثل الكلمة الإنجليزية "Heli" التى تحل محل كلمة "Helicopter" أو كلمة "Helium" . . . إلخ .

(٣) **التجاور** : Proxinty ، حيث يستطيع المستخدم البحث عن الكلمات أو العبارات التى تتجاور مع بعضها البعض ، ويجب أن تتلاحق وتتجاور الكلمات الواردة وتحدث فى نطاق محدود من الكلمات . وتمثل أحد الأوجه الذى يحدد ما إن كان البحث المتجاور ذا اتجاه ثنائى أو متعدد "Uni - or bi - directional" .

(٤) **الجدال المتعدد** : Multiple Arguments ، يوضح هذا الأسلوب القدرة على تحديد الجدال المتعدد الخاص بالبحث باستخدام مشغلات المنطق البولينى Boolean Logic التى تتمثل فى التالى : أداة العطف «و AND» ، أداة الإختيار «أو OR» ، أداة النفى «لا NOT» . ويمكن ربط وتشبيك هذه الأدوات معًا .

(٥) **بحث العبارة** : Phrase Search ، فى مقدرة هذا الأسلوب تحديد عبارة ترتبط بأغراض البحث مثل عبارة «انتهاء المكتبة اليوم The Library Ends Today» حيث يمكن أن يسهم استخدام «كلمات التوقف» السابق الإشارة إليها فى تحديد كمية البحث المرتبطة باستخدام العبارة المعينة .

(٦) **البحث الغامض** : Fuzzy Search ، يمكن استخدام أساليب البحث الغامض فى مراعاة التهجية الخطأ Misspelling للكلمات ، أو الأخطاء المرتبطة بالتعرف الضوئى

على الحروف OCR المتواجدة فى الوثائق . على سبيل المثال ، البحث الغامض عن كلمة مثل «علاقى Relational» يجب أن يعثر على هذه الكلمة فى أثناء عملية البحث وتحديد ارتباطاتها ، كما يمكن ملاحظة إتجاه البحث الغامض إلى التوسع فى عدد الوثائق المسترجعة بدلاً من تقليلها.

(٧) **البحث عن التركيب** : Search on Structure ، يمكن أن يبحث المستخدم عن الكلمات أو العبارات المتواجدة فى تركيب الوثيقة ، وعلى سبيل المثال ، يعرف محرك بحث النص المفتوح Open Text's Text Search تركيب أو هيكل الوثائق التى تستخدم لغات برمجة متقدمة مثل لغة HTML ، أو لغة SGML الخاصة بإعداد مواقع الويب على الإنترنت.

(٨) **التساؤل بواسطة المثال** : Ouery - by - Example ، تسمح بعض منتجات برمجيات قاعدة بيانات النص مثل أداة خادم بحث فولكروم Fulcrum's Search Server بأن يأخذ المستخدم وثيقة ما ويستخدمها كمثال لتحسين بحث إضافى ،مثل استخدام أمر «ابحث عن وثائق أكثر من هذه الوثيقة» .

(٩) **وضع نتائج البحث فى المضمون** : فى العادة ، ترتب منتجات برمجيات بحث النص الكامل الوثائق المسترجعة وتعرضها فى ترتيب تنازلى طبقًا لدرجة التوافق مع معايير التساؤل . ويستخدم كل منتج الألجوريثم الخاص به لتقدير التوافق بناء على عدد الكلمات وتتابعها وأى معايير أخرى . وفيما يتصل بمضمون الوثيقة ، يرتبط ذلك بالقدرة على تجميع أو تسلسل الوثائق بواسطة الموضوع ذاته .

٤ - مرونة الإنتشار : Flexibility of Deployment

يحتم على منتجات برمجيات بحث النص أن تكون مرنة حتى تتفق مع استخدامات كثير من المكتبات ومراكز المعلومات ، من خلال الطرق التالية :

(١) **مساندة اللغات المتعددة** : Multiple Language Support تساند بعض البرمجيات المتاحة مثل : Fulucm's Serarch Server ، Open Text's Text Search ، و Verity's Topic إمكانية بحث النص الكامل للوثائق فى لغات متعددة . كما يساند

منتج IBM's Text Extending المرتبط بقاعدة بيانات DB2 الخاصة بها بحث النص الكامل فى سبعة عشر لغة من بينها اللغة العربية .

(٢) **أشكال النص المتعددة** : Multiple Text Formats تساند معظم منتجات برمجيات النص الكامل أشكال الوثائق المتعددة فيما يتصل بالتكشيف والعرض . وتتمثل الخاصية الرئيسية المرتبطة بذلك فى تضمين كل الأشكال فى كشاف واحد .

(٣) **القدرات الموزعة** : Distributed Capabilities ، حيث أن الوثائق تنتج فى الغالب محليا على قمة المكتب Desktop ، لذلك يوجد كثير من التوجهات أو القضايا التى ترتبط بالمشاركة فى الوثائق وإمكانية إتاحتها عبر شبكة الإنترنت . وفى هذا الإطار توجد عدة أسئلة ، تحدد الإجابة عليها القدرات الموزعة المطلوب تواجدها :

- هل يجب أن يحدد موقع تواجد نسخة الوثيقة ؟
- هل يجب أن يكشف محرك بحث النص الوثيقة ويجعل فى الإمكان الوصول إليها مباشرة ؟
- هل فى الإمكان تكشيف الوثيقة فى إطار الشكل المتواجد محليا ؟
- هل يمكن التوسع فى البحث لقواعد بيانات الوثائق المتعددة ؟
- .... إلخ .

(٤) **مساندة أنواع البيانات الأخرى** : Support to Other Data Types تساند معظم منتجات بحث قواعد بيانات النص الحقول الهيكلية ، ويقدم بعض هذه المنتجات مثل DB/Text Works إمكانية الوصول إلى بيانات الوسائل المتعددة التى ترتبط مع الوثيقة ، ولكنها لا تعبر عن محتوى مبنى على استرجاع بيانات الوسائل أو الوسائط المتعددة .

٥ - فهم معنى النص : Undrstanding The Meaning of Text

فى الواقع ، يتضمن التحليل اللغوى أو معالجة لغة النص الطبيعية تحديد وتفسير الجمل لتقرير موضوع النص أى التعرف على معناه وتقرير المفاهيم الأخرى التى يغطيها بغض

النظر عن الكلمات المستخدمة فقط ، حيث أن الكلمة نفسها يمكن أن يكون لها أكثر من معنى مختلف ومستويات مختلفة من الأهمية اعتمادًا على كيف وأين تستخدم هذه الكلمة فى سياق جملة أو وثيقة .

وتشتمل بعض مزايا التحليل اللغوى على ما يلى :

(١) **تساؤلات اللغة الطبيعية** : Natural Language Queries ، التى تسمح للمستخدم من إنشاء استفسارات أكثر من أشكال حرة غير مقيدة . ويفسر محرك بحث النص ما يحتاجه المستخدم ويريده ، كما يوفر مرونة وسهولة الاستخدام .

(٢) **مستخلصات الوثيقة** : Document Abstracts ، يمكن أن تشتمل بعض منتجات بحث النص مثل Oracle's Text Server التى ترتبط بمضمون البحث على توفير مستخلص أو ملخص لمحتوى الوثيقة الذى يظهر آليا ، كما قد تتضمن بعض المنتجات خاصية سرعة القراءة عند تصفح بيانات النص . وقد أضافت أداة شركة Verity قدرة الاستخلاص فيما يتصل بمحرك Topic الخاص . ومن الملاحظ أن المستخلصات تسمح للمستخدم إقرار مدى إمكانية استرجاع كل نص الوثيقة أو الاكتفاء بها ، مما قد يوفر سعة نطاق الشبكة إلى حد كبير .

٦ - تبسيط تطوير التطبيق : Simplifying Application Development

تتضمن معظم منتجات بحث النص كما فى حالة أدوات كل من Oracle ، Verity ، و Excalibur خواص واجهات التفاعل ولغة الاستفسار لاسترجاع النص . وفى بعض المنتجات الأخرى كما فى حالة أدوات Fulcrum ، Informix ، IBM توجد خواص أخرى ترتبط بمدخل لغة الاستفسار الهيكلية SQL الممتدة للوصول إلى النص .

٧ - تكامل النص فى نظام إدارة قاعدة البيانات العلاقية :

حتى يمكن تضمين النص مع البيانات الهيكلية لقاعدة البيانات يجب تضمين مجموعة من المتطلبات المعيارية الخاصة بذلك ، منها : التوسع فى أى نظام إدارة قاعدة بيانات علاقية لكى يتعامل مع الأشياء المعقدة بفعالية ، ويعتبر ذلك عملية معقدة حيث يشتمل على طرق تكشيف واسترجاع معقدة مبنية على محتوى النص الذى يوصف لتداول هذا المحتوى

الموضوعى مباشرة ؛ تنفيذ طرق مختلفة للوصول إلى البيانات التى تخزن أدوات إدارة أو مديرى بيانات غير متجانسين Heterogeneous ؛ تطوير طرق الوصول لما قد يطلب من أنواع البيانات غير المتجانسة . وتعتبر هذه الخواص توسعات فى منتجات بحث النص التى تخدم إدارة قاعدة البيانات حتى تلبى حاجات المستخدمين فى الوصول إلى البيانات المتكاملة التى تصبح معلومات مفيدة .

مما سبق ، يتضح أن مفتاح نجاح أى تكنولوجيا ترتبط بقدرات بحث النص يتمثل فى سهولتها حتى يمكن استخدامها والتوصل السريع إليها ، مما يحتم ضرورة توافر أدوات الحاسب العميل الرسومية المرتبطة بإدارة عملية بحث النص . على سبيل المثال ، تشتمل أداة Oracle على مجموعة تسهيلات إدارية رسومية ، كما تتضمن على بعض أدوات بحث النص المرتبطة بإدارة الوثائق وخادمات الويب . . . إلخ .

نظم إدارة الوثائق

تعتبر نظم إدارة الوثائق تطبيقات تدير الوثائق وتراقبها ، حيث يقوم المستخدم بتخصيص قيم حقل من حقول قاعدة البيانات (المؤلف ، تاريخ النشر ، الكلمات الرئيسية ... إلخ) لوثيقة معينة . وتعرض هذه الحقول خواص مجموعة وثائق تحفظ فى قاعدة البيانات ، إما كقاعدة بيانات على قمة المكتب Desktop أو نظام إدارة قاعدة بيانات علاقية RDBMS. وفى العادة ، لا تحول إدارة الوثيقة فى عنصر بيانات ولكنها تنشئ سجل قاعدة بيانات مرتبطة بالوثيقة .

ويعتبر بحث النص كله خيارا منفصلاً فى معظم نظم إدارة الوثائق مصحوبا بالقدرة على البحث عن خواص الوثيقة أو النص كله أو كليهما . وحيث أن النص كله يضيف أعباءً إلى النظام ، فقد لا يريد المستخدمون تضمين كل الوثائق فى كشاف النص الكامل . وتسمح بعض النظم للمستخدم أن يحدد خواص معينة لتقدير أى الوثائق يجب أن تكشف فى النص الكامل . وتوفر معظم برمجيات قواعد البيانات محركات بحث النص ، كما فى حالة Verity ، Fulcrum ، Excallibur ... إلخ . بالإضافة إلى الاسترجاع بواسطة الخواص وبحث كل النص ، تقدم معظم نظم إدارة الوثائق خواص إدارة الوثائق والرقابة عليها ، وتشتمل هذه الخواص على الفحص الداخلى والخارجى لرقابة التكرارات ، الإصدار ، الأمن ، التخزين والخدمات الأخرى .

ويعتبر تدفق العمل Workflow تكنولوجيا مستقلة ، بدأت كثير من نظم إدارة الوثائق من استخدامها للتكامل . وصارت كثير من البرمجيات المتاحة توفرها لنظم إدارة قواعد البيانات العلاقية RDBMS مثل قواعد بيانات Oracle ، Sybase ، Informix ... إلخ . وعلى سبيل المثال ارتبط نظام النص المفتوح Open Text مع نظام Odesta الذى يجعل منتجات بحث النص متكاملة مع قدرات نظم إدارة الوثائق وتدفق العمل . وتقدم بعض البرمجيات مثل Lotus Notes حزمة برمجيات تؤدى إلى تعاون المجموعة للمشاركة فى المعلومات على شكل الوثائق . وعلى الرغم ، من أن هذه الحزمة لا تمثل وثيقة معينة أو نظام إدارة قاعدة بيانات إلا أن كثيرا من الهيئات المستخدمة لها تبنى بعض التطبيقات التى ترتكز على النص المعقد المتوافر فى هذه البرمجيات التى تخزن معلومات الوثيقة فى شكل

مجموعة حقول Fields التى تكون فى إطار حقل هيكلى مثل حقل التاريخ ، الاسم ، العنوان . . . إلخ ، أو حقل غير هيكلى Unstructured يفسر كنص مع تهيئة قليلة مرتبطة بالألوان ، الأنماط ، الجداول الضمنية ، الصوتيات أو الفيديو الضمنى . . . إلخ . كما تشتمل هذه البرمجيات على محرك بحث النص الكامل كما فى حالة برمجيات Verity . وفى هذه البرمجيات تعرف كل قاعدة بيانات لتكشيف النص الكامل ، ويقارن ذلك بالتكشيف على مستوى الجدول فى نظام إدارة قاعدة بيانات علاقية RDBMS ، كما يخزن كشاف النص الكامل ، وبذلك تتم الاستفسارات عندما تتوافر قاعدة بيانات النص حيثما تتواجد قاعدة البيانات التى إما أن تكون متاحة ومخزنة فى الحاسب العميل أو الحاسب الخادم .

وفى الماضى ، أهمل مقدموا نظم إدارة قواعد البيانات العلاقية RDMS كل أنواع البيانات المعقدة إلى حد كبير . وما تقدمه قاعدة البيانات العلاقية يتمثل فى نوع البيانات BLOB الذى يتضمن فى ذاكرته ما يقرب من ٢ جيجا بايت من البيانات المخزنة فى ذاكرته المصمم للتعامل مع أنواع البيانات الرقمية والهجائية ذات الطول الثابت والمتغير المتضمن المحتويات التاريخية والزمنية ، وقد يشتمل ذلك على فهم بسيط لتركيب نوع البيانات BLOB الداخلية ، كما أنه لا يفسر المحتويات الخاصة بذلك . ومن البدائل المتاحة إعداد بيانات تعليم بيانات BLOB مصحوبا بكلمات رئيسية Key words تخزن فى حقول أخرى متاحة على قاعدة البيانات التى تؤخذ من تحرير البيانات فى نظام الملف ، أو ترك معالجة البيانات BLOB بدلا من تخزين واسترجاع تطبيق نهاية المقدمة Front - end .

وفى الوقت الحاضر ، توجد كثير من التطورات الجوهرية والأساسية التى يوفرها مطوروا نظم إدارة قواعد البيانات العلاقية ، حيث يمدون خادم قاعدة البيانات العلاقية بإمكانيات التداول لأنواع البيانات المعقدة غير المقتصرة على النص فقط . وتعتبر هذه التوسعات غير هامشية ، وتتطلب فى العادة تضمين مكونات ثرية فى نطاق إدارة قاعدة البيانات العلاقية .

وبذلك تصبح هذه النظم الحديثة الأساس الجوهرى للوصول المتكامل للمعلومات عبر المكتبات ومراكز المعلومات الكبيرة . وأصبحت هذه النظم توفر إمكانيات قواعد البيانات المفتوحة التى تتيح مستويات مرونة أعلى للمستخدمين .

وقد اتبعت شركة ميكروسوفت مدخلا مختلفًا عما سبق فيما يتصل بمساندة البيانات غير التقليدية من خلال نظام التشغيل الخاص بالنوافذ Windows الذى يقوم بتشغيل وإدارة قواعد البيانات عليه . ويتميز ذلك بتضمين واجهات التفاعل Interfaces لكل أنواع البيانات فى نطاق النظام المبنى على الشيء Object - oriented بدلاً من بناء معمارية مستقلة فى نطاق نظام إدارة قاعدة البيانات الذى يحذف أو يستبعد الحاجة لإعادة اختراع نظام جديد، مما يجعل واجهات التفاعل أكثر ملاءمة للتطبيقات والخدمات ، ويتوقع الوصول إلى أنواع البيانات الممتدة فى شكل OLEDB كنوع جديد من واجهات التفاعل للوصول إلى البيانات.

وتعتبر نظم إدارة قاعدة البيانات العلاقية الشيء ORDBMS مدخلا جديداً نسبيًا لمشكلة تداول البيانات المعقدة فى قاعدة البيانات بفعالية أكبر . وتتجه هذه النظم المرتبطة بالتوضيحات والرسومات النابعة من تكنولوجيات المعلومات الرسومية إلى تقديم المعلومات الأحسن النابعة من الواقع العلاقى والموجه نحو الشيء فى آن واحد . وتبنى هذه النظم على أساس علاقى يرتبط بالأمن والسلامة والمصداقية ومساندة لغات التساؤل أو الاستفسار الهيكلية SQL التى ترتبط بها ، بينما يمكن إضافة إمكانية التوسع والامتداد الموجه نحو الشيء لتداول البيانات المعقدة . ويعمل هذا المدخل التطورى على رفع المهارات وزيادة المعرفة الحالية التى اكتسبت فى الواقع العلاقى بينما تعمل على تطوير أنواع جديدة من التطبيقات .

وتشتمل نظم إدارة قاعدة البيانات العلاقية الشيء ORDBMS على مزايا تفوق إلى حد كبير نظم إدارة قاعدة البيانات العلاقية فحسب فيما يختص بالبيانات المعقدة ، مما يتيح إمكانيات أكبر لتسويقها مستفيدة فى نفس الوقت بهيكلية قاعدة البيانات العلاقية ، وتساند المعالج المتعدد Multiprocessor ،والمعالجة المتوازيه ، والبيانات الموزعة وأدوات الإدارة والمراقبة المعقدة .

ويشتمل نظام إدارة الوثائق على عدة مراحل خلال دورة حياة النظام . ويمكن تعريف مهام عديدة للمراحل المختلفة . ومن المهم وضع القواعد والمعايير لتأكيد التنسيق بين كل شخص متضمن فى هذه المهام .

وفيما يلى وصف مختصر لمراحل نظام إدارة الوثائق :

١- مرحلة التصميم : The Design Stage

(١) **الفريق الاستراتيجى** : يمكن الحصول على النتائج الأحسن عن طريق توفير فريق عمل من تخصصات متداخلة لتفسير استراتيجية المعلومات الإلكترونية ومراقبتها . كما أن الطريقة الأحسن لتعظيم إدارة المعلومات الإلكترونية تتمثل فى تفسير الاستراتيجية الشاملة المتماسكة التى تؤكد أن كل فرد مختص متضمن فى ذلك . وتشتمل مسئوليات فريق استراتيجية المعلومات الإلكترونية على التالى :

- تحديد متطلبات المستخدمين (مشتملة على الأوجه القانونية) .
- تعريف السجلات المهمة .
- تفسير القواعد لخطة تصنيف كفء .
- تفسير المعايير والمواصفات لتأكيد استقلالية البيانات من الوسائل لضمان تحميلها .
- تحديد خطة التقويم والمراجعة .
- تعريف المسئولين عن كل مهمة .
- تحديد سياسة التدريب والتوعية فى المكتبة أو مركز المعلومات المختص .
- مراقبة تنفيذ النظم الجديدة .

(٢) **الأوجه القانونية** : تعتبر إحدى القضايا الرئيسية التى يجب مراعاتها فى استراتيجية المعلومات الإلكترونية ، وتتضمن التالى :

- قيمة المعلومات الإلكترونية القانونية .
- قد يتنوع ويختلف معانى الألفاظ من دولة لأخرى .

٢- إنشاء المعلومات الإلكترونية : Creating Electronic Information

من المهم اعتبار كل المهام المتضمنة فى نظام إدارة الوثائق من البداية . وعلى ذلك يمكن تطبيق القواعد المشتركة عند إنشاء المعلومات الإلكترونية . وسوف يسهل ذلك الاتصال بين المسئولين لمعالجة المعلومات فيما بعد . وتتضمن القواعد التالية عند إنشاء المعلومات الإلكترونية :

- يجب أن تعرف كل وثيقة أو قاعدة بيانات بوضوح بواسطة المكتبة أو مراكز المعلومات المسئول عن إدارة المعلومات الإلكترونية .
- يجب تقديم التوثيق للمعلومات الإلكترونية الذى يحفظ بواسطة الشخص المسئول عن إدارة المعلومات الإلكترونية فى كل مرحلة .
- يجب إنشاء الإجراءات لمعالجة المعلومات الإلكترونية غير المتأكد من مصدرها .
- لا يجب التخلص أو تغيير البيانات بدون الموافقة الصريحة بذلك .

وفيما يتصل بالمعلومات الإلكترونية أو الوثيقة الورقية ، يمكن ملاحظة التالى :

- فى العادة تحفظ المعلومات على الورق فى شكل ورقى . وقد تساعد أدوات البحث الإلكترونية للإدارة والاسترجاع .
- تحفظ المعلومات الإلكترونية على وسائط أو وسائل إلكترونية ، ويمكن مسح Scan الوثائق الورقية أو بعض المعلومات النصية التى تشتمل عليها للحصول على السجلات الإلكترونية لكى تبسط البحث والاستشارة .

وتشتمل مرحلة إنشاء أو خلق المعلومات الإلكترونية على المهام التالية :

(١) تكامل المعلومات : Integration of Information

تعتبر عملية إعادة تنظيم البيانات واختيارها لإنشاء بيانات مدمجة أكبر مهمة أساسية، حيث يجب حفظ حق خصوصية المعلومات . وفى بعض الأحيان قد يكون من الضرورى دمج المعلومات من المكتبات أو مراكز المعلومات العديد .

وعندما تنشأ وثيقة من قاعدة البيانات ، توجد حلول كثيرة للتخزين الطويل الأجل لها ، منها :

- تجسيد قاعدة البيانات باستفسار معين لإنتاج وثيقة تقليدية ، وتحديد أى حالة تفقد بعض التجميعات الممكنة .
- إذا كان للتطبيق ذيل مراجعة كامل ، فإن قاعدة البيانات سوف تصدر عندما يكون النظام خارج التشغيل .

- حفظ قاعدة البيانات والتطبيق الذى ينشئ الوثيقة بطريقة ديناميكية .

ويمكن أن يتنوع حجم وحدة المعلومات اعتمادًا على الإختيارات المتخذة . ومن جهة ، يعنى العدد الضخم من المراجع ووصلات النص المتشعب Hyretext بين الوثائق حاليًا ، توفير البيانات على الإنترنت مثلا لتشكيل سجل واحد ، ومن جهة أخرى ، يجب أن تكون وحدات المعلومات بحجم معقول لكى تصبح ممكنة الاستخدام .

(٢) تحويل المعلومات : Conversion of Information

يوجد حلان رئيسيان لتحويل وثيقة من الشكل الورقى إلى الشكل الرقمى :

- مسح الوثيقة بسهولة للحصول على شكل لها .
- مسح الوثيقة وتكويدها فى شكل إلكترونى باستخدام التعرف الضوئى على الحرف OCR ، أو من خلال الرسومات Graphics ، كما يحدده الشكل التالى :

شكل (٦ - ١) : التحويل من وثيقة ورقية إلى وثيقة إلكترونية

ويتضمن أحد الخيارات الأخرى تجميع هذين الاتجاهين ، أى توفير الأشكال الخام ثم استخدام التعرف الضوئى على الأجزاء المشتملة على النص .

ويشتمل نوع التحويل على تغيير أحد الأشكال الرقمية فى شكل رقمى آخر . ويجب أن يعمل ذلك فقط لتحويل سجل قائم فى شكل معين إلى شكل أكثر معيارية وأكثر دواما ، أو للتحويل فى معيار سوف يسمح بإمكانيات أكثر . ويمكن أيضًا استخدام مرحلة ثالثة فى تحويل وثيقة ورقية للحصول على سجل إلكترونى بشكل مهيكل بدرجة عالية ، كما فى حالة وثيقة بهيكل ظاهر أو قاعدة بيانات .

وعلى الرغم من أن تحويل السجلات الورقية إلى الشكل الرقمى يتضمن وثائق فى الأساس ، فقد يكون من الممكن ، فى بعض الأحيان ، إعطاء هيكل قاعدة البيانات للسجل الإلكترونى باستخدام هذه المرحلة ، كما فى الشكل التالى :

شكل (٦ - ٢) : تحويل الشكل الرقمى

(٣) التحول من الورق أو الميكروفيلم إلى الشكل الممسوح :

From Paper or Microfilm to Scanning Image

يساعد مسح Scanning وثيقة توفير محتوى الصفحات الورقية للملفات كمبيوتر مشتملا على شكل الوثيقة المبدئى فى درجة وضوح Resolution معينة ، وتقاس درجة وضوح الماسح Scanner فى نقاط للبوصة dots per inch (dpi) . وفى الوقت الحالى يمكن أن تحقق الماسحات درجة وضوح تتمثل فى ٣٠٠ أو ٦٠٠ نقطة فى البوصة بالألوان .

وعلى سبيل المثال ، يشتمل جهاز الفاكس على ماسح ، نظام لإرسال البيانات عبر خط التليفون ، وطابعة عند المستقبل ، وتتراوح درجة وضوح الماسح للفاكس ما بين ١٠٠ إلى ٢٠٠ نقطة فى البوصة فقط .

وفى هذا الصدد يحتاج إلى ملاحظة عدد من القواعد الأساسية عند مسح الوثائق ، حيث أن جودة الناتج تعتمد على ذلك إلى حد كبير . ولا يجب تخزين النص كملف شكل، إن لم يكون ذلك قصيرًا أو أن يكون التعرف الضوئى على الحروف OCR غير ممكن فى حالة التوقيع مثلا ، وفى حالات أخرى يكون من المفضل تكوين النص . ومن قواعد مسح الوثيقة ، ما يلى :

- كل الوحدات المحتاجة لسجل مبدئى يجب أن تنقل على الوسيلة أو الوسيط .
- الوصلات بين السجل وباقى نظام إدارة الوثائق كالأرشيف يجب أن تحفظ .
- جودة الوثيقة الأصلية يجب أن تكون جيدة لتأكيد أقصى جودة ممكنة للمخرج المطبوع بعد التحويل إلى الشكل الرقمى .
- قبل مسح الوثيقة ، يجب إختبار عينة .

- حتى ولو أنجز المسح خارج المكتبة أو مركز المعلومات ، فإن المنظمة التى تؤدى العمل يجب أن تختبر الوثائق الرقمية للجودة والإكتمال .
- يجب تصميم النماذج لتحويلها الممكن مع مراعاة الشكل الإلكترونى فيما يتصل بالأبناط والأحجام ومواقع الحقول . . . إلخ .

(٤) التحول من الأشكال الممسوحة إلى الشكل المكود المرمز :

From Scanned Image to Encoded Format

يساعد التعرف الضوئى على الحروف OCR الكمبيوتر فى قراءة النص . وتعمل أو تشغل برمجيات التعرف الضوئى على الحروف من ملف يشتمل على شكل أو صورة النص المراد قراءته ، على سبيل المثال ملف منشأ بواسطة الماسح ، وبرمجيات التعرف الضوئى تحلل ملامح الحروف وتنتج ملف فى شكل نص يمكن تحريره بواسطة برنامج معالج الكلمات WP .

ولا يكون التعرف متقنا ، عندما لا يقدر الكمبيوتر تعريف حرف ، وتعليمه للمشغل البشرى لكى يتعرف عليه . ومن النادر جداً ، أن يقرأ الكمبيوتر الحرف بطريقة غير صحيحة . ويمكن أن تساعد أدوات التصحيح الآلية فى تعظيم التصحيح . ولكن يجب أن يقوم المشغل البشرى بعمل اختبار أو فحص بعد التعرف الضوئى على الحروف . وقد يشتمل هذا الفحص أيضًا على استخدام أدوات برمجيات لذلك .

وبعد مسح وثيقة متضمنة رسومات ، يمكن استخلاص النص باستخدام التعرف الضوئى على الحروف . ويجعل ذلك فى الإمكان تحرير النص ، استخدام أجزاء منه أو تكشيفه لاستشارة أسهل .

(٥) التحول من أحد الأشكال الرقمية إلى أخرى :

From One Digital Format to Another

توجد حالتان حيث يكون فيهما من المفيد استنساخ المعلومات من أحد الأشكال الرقمية إلى أشكال أخرى وهما :

- الهجرة من شكل إلى شكل أكثر متانة لجعل الإنقرائية الطويلة الأجل والاستشارة أسهل .

- إضافة هيكل لملف سطحى (نص خالص) للحصول على وثيقة هيكلية أو قاعدة بيانات تكون أسهل للاستشارة .

وعندما تحول المعلومات الإلكترونية من أحد الأشكال إلى شكل آخر ، يجب العناية بتجنب فقد البيانات العرضى .

(٦) تقويم المعلومات الإلكترونية : Appraisal of Electronic Documents

لا يخدم حفظ البيانات أى غرض إن لم يمكن استشارته عند الطلب . وعندما تلتقط المعلومات كسجل لتوضيح دليل نشاط وجعل التقويم نقطة مهمة بصفة معينة يجب تقويم صحة وسلامة وقيمة الوثيقة بعناية .

ومن مهام التقويم ما يلى:

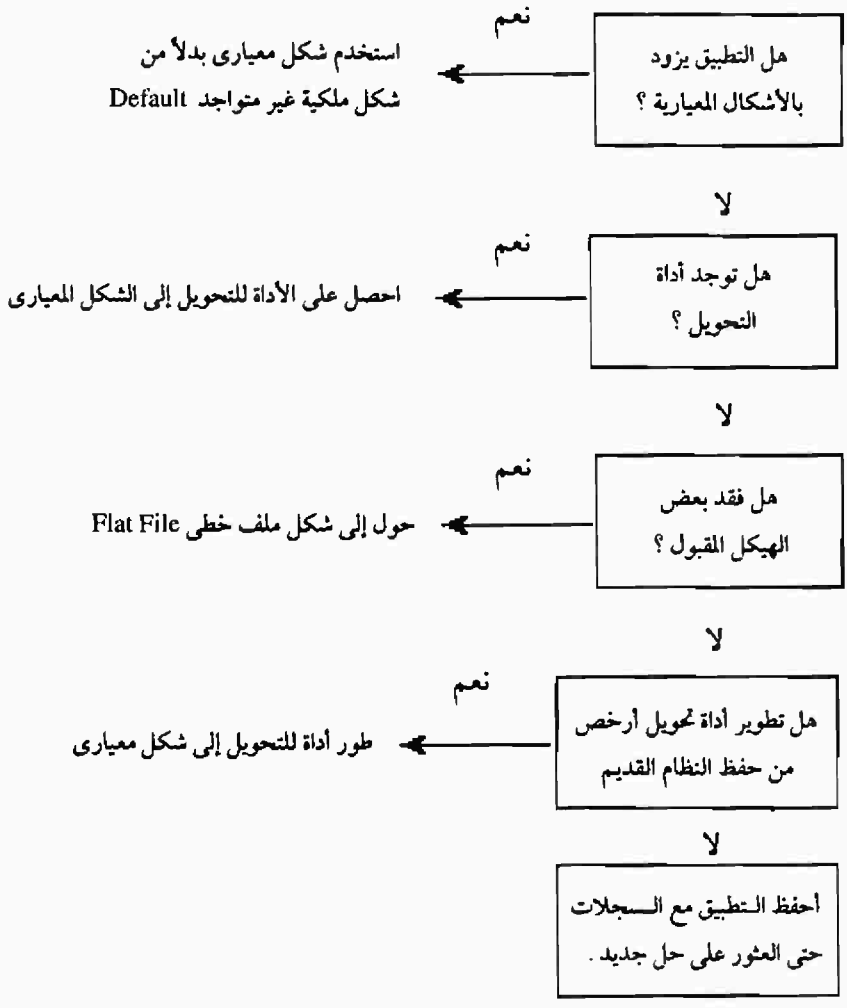
أ - فحص السجلات : تنبع معظم المشكلات خلال تحويل المعلومات وإرسالها .

ب - استبعاد السجلات : التى لا يوجد لها استخدام أو قيمة معينة .

جـ - تفسير الأداء الأحسن أو غير السليم للاستبعاد بوسطة المكتبات ومراكز المعلومات .

د - تحديد خيارات أشكال الملكية Proprietary Formats التى تتمثل فى الشكل التالى:

شكل (٦ - ٣) : كيفية التعامل مع أشكال الملكية المعيارية



(٧) إدارة المعلومات وتصنيفها :

Management and Classification of Electronic Information

الهدف من إدارة المعلومات الإلكترونية حفظ مصداقية المعلومات وموثوقيتها وسلامتها وإمكانية اختبارها عبر الزمن . ويتطلب ذلك أن يفسر مضمون المعلومات جيداً . وعندما يكون المحتوى السياق والهيكل كاف لتحديد إثبات نشاط ما عندئذ تصبح المعلومات سجلا .

تتضمن إدارة مراحل نظام إدارة المعلومات الإلكترونية أيضًا مهاما أخرى . وقد تنقل المسئولية لإدارة مجموعة معلومات إلكترونية معينة لمكتبة أو مركز معلومات آخر .

وتشتمل إدارة المعلومات الإلكترونية على المهام التالية :

- **تسجيل مقتنيات المعلومات الإلكترونية** ، ويتطلب ذلك تحديث تعقب مراجعة Audit Trail البيانات فى الاستفسار المعين .
- **التكليف** عند تضمين مكتبات أو مراكز معلومات عديدة ، يجب أن يكون التكليف أو التخصيص للمكتبة أو مركز المعلومات الصحيح كما يجب تعيين مدير معلومات إلكترونية .
- **المتابعة** تتضمن تنسيق المراحل العديدة فى المعالجة المكونة من استلام المعلومات الإلكترونية ، التحويل ، الحفظ ، الاستخدام ، والنقل لمكتبة أخرى .
- التصنيف يصبح أسهل فى استرجاع معلومات إلكترونية معينة .
- **القرار** للنقل إلى مكتبة أو مركز معلومات أو مستخدم آخر .

وتتضمن إدارة المعلومات الإلكترونية مجموعات أشخاص أكبر . وقد تكون أدوات تدفق العمل Workflow عاملاً مساعدًا لإدارة التبادل بين المجموعة المتعاونة .

وأحد المهام الأكثر أهمية ومن المحتمل أن تكون الأكثر تعقيدًا هى تصنيف المعلومات . ويجب أن يكون نظام الترميز واضح بدرجة كافية لكى يمكن فهمه من قبل المكتبات أو المستخدمين الآخرين .

ويساعد هيكل المعلومات الإلكترونية تحديد موقع وحدة معلومات معينة فى وثيقة أو قاعدة بيانات ، بينما يساعد تصنيف المعلومات تحديد موقعها بين كل المعلومات المختزنة .

ويجب أن تستخدم خطة تصنيف عامة مهما كان الدعم المقدم للمعلومات . ويمكن إضافة معايير إضافية لتكشيف المعلومات لكى تراعى خصوصية المعلومات الإكترونية . ومن المعايير المرتبطة بخطة تصنيف المعلومات الإلكترونية ، ما يلى :

- نوع الوثيقة .

- التواريخ (الإنتاج ، الانتهاء) .
- المؤلف أو المؤلفون الأفراد والهيئات .
- الموقعون Signatory على الوثيقة .
- وجهة Destination الوثيقة الإلكترونية للأشخاص أو المؤسسات .
- النسخ للأفراد والهيئات .
- رقم المعلومات الإلكترونية وإصداراتها .
- التكليف (التاريخ ، الملف، المكتبة ، . . . إلخ) .
- الموضوع .
- المشروع أو النشاط .
- الكلمات الرئيسية Keywords .
- اللغة .
- عدد الصفحات .
- وضعية أو حالة المعلومات الإلكترونية رسمية أو غير رسمية ودرجة السرية .
- الوثائق المرتبطة والوصلات مع الوثائق الأخرى .
- أوجه أخرى تعرَّف بواسطة المستخدم .

(٨) نقل المعلومات الإلكترونية : Transfer

فى نهاية الجزء النشط من نظام إدارة الوثائق ، يمكن إرسال السجلات إلى المكتبة أو الأرشيف الخاص ، علما بأن كل السجلات ليس لها قيمة حفظ أرشيفية ، لذلك يمكن استبعاد الوثائق التى ليس لها استخدام أو قيمة معينة فيما بعد .

ويوجد نوعان رئيسيان لإرسال المعلومات :

- إرسال المعلومات الإلكترونية طبيعيًا .
- نقل المسئولية .

وفى العادة ، تنقل المعلومات فى شكل سجل بعدما يحصل عليه كسجل . ويمكن إرسالها لمكتبة أو مستخدم آخر أو نقلها إلى خدمات أرشيفية .

ولا تحتاج مسئولية المعلومات الإلكترونية وكفايتها الطبيعية إلى أن تتوافق بالضرورة ، ويمكن حفظ المعلومات الإلكترونية بواسطة المكتبة أو مركز المعلومات أو المستخدم . كما يمكن أن تؤدى سهولة نسخ المعلومات الإلكترونية أو نقلها إلى خلق مسئولية فصل الدورين عن بعضهما البعض .

على أى حال ، يجب القيام بالفحوص والإختبارات دائمًا قبل نقل المعلومات الإلكترونية التى تتضمن التالى :

- عمل نسختان من البيانات .
- مقارنة البيانات مع التوثيق المستمد من المورد باستخدام برنامج إحصائى .
- تعريف وتوثيق أى أخطاء فى التوثيق .
- عمل اختبارات وفحوصات متداخلة مثل فحوصات فى السجل أو الفحوصات المتغيرة لتحقيق ثبات البيانات .
- استشارة مورد المعلومات الإلكترونية عند حدوث أى مشكلة أو صعوبة فى تعريف الأكواد أو إذا وجدت أخطاء وعدم توافق فى البيانات .
- توثيق الملفات الطبيعية مع بيان أى صعوبات يمكن التغلب عليها .

تكنولوجيا قواعد البيانات المتطورة

(١) المقدمة :

حاليًا ، صارت قواعد البيانات تعتمد على شبكة الإنترنت وشبكات الإنترانت الواسعة الإنتشار على الصعيد الدولى . ويقدم معظم مطورى قواعد البيانات ومتعهديها الرئيسيين منتجاتهم من البرمجيات التى صممت بهدف مساعدة المستخدمين فى نشر بياناتهم على شبكة الويب التى تستخدم من قبل المستخدم النهائى كأداة للاستفسار فى قواعد البيانات المحملة عليها . وفى كثير من الأحيان ، أصبح المديرون على كافة توجهاتهم وتخصصاتهم يقومون بنقل كم أكبر من البيانات إلى المستخدمين النهائيين حتى يكتسبوا الكفاءة المطلوبة إلى جانب توفير وقتهم وجهدهم فى تداول كثير من الطلبات التى تتعلق بتحليل البيانات وإنتاج التقارير بأنفسهم .

وقد صارت خواص ترابط وتواصل شبكة الإنترنت الدولية مع البرامج الوسيطة Middleware تؤثر على استراتيجيات وسياسات مطورى وموردى قواعد البيانات كما فى حالات برمجيات قواعد البيانات مثل : M.S. SQL Server ، IBM DB-2 ، Oracle، Sybase ، Informix ، . . . إلخ . ومن العوامل الرئيسية فى تخزين قواعد بيانات النص أو قواعد بيانات التوجه الشيئى والوصول إليها ارتباطها باستخدام نموذج قاعدة بيانات الشكل العلاقى أو الهجين Hybrid الذى يختص بتوسيع التطبيقات مع قواعد البيانات المفسرة من قبل ، كما بدأ كثير من المطورين والموردين تضمين مستودعات النصوص والأشكال فى منتجات قواعد بياناتهم .

(٢) مطوروا قواعد بيانات النص والأشكال :

تتعدد الأساليب المستخدمة فى استرجاع البيانات من قواعد بيانات النص التى تتعامل بصفة متزايدة مع شبكة الإنترنت وشبكات الإنترانت التى تتوافر للمكتبات ومراكز المعلومات.

وفيما يتصل بصناعة إنتاج قواعد بيانات النص بالإضافة إلى الأشكال والرسومات يمكن تحديد اثنى عشر مطورًا أساسيًا فى ذلك ، كما يلى :

1 - Informix Software Inc.

- Menlo Park, California .
- DB : Illustra's Data Blade

2 - Microsoft Corporation.

Redmond, Washington

- Windows NT, 2000 .
- BB : SQL Server

3 - Oracle Corporation

Redwood Shores, California

- Network Computing Architecture (NCA)
- DB : Oracle

4 - IBM Corporation

Armonk, New York

- DB : DB2; Universal database

5 - Sun Microsystems

Mountain View, California

- DB : Object Database Management Group (O D B M G)

6 - RedBrick Systems

Los Gatos, Califorina

- Online database with distributed technology and online complex processing (OLCP)
- DB : Red Brick Warehouse.

7 - Verity Inc.

Sunnyvale, California

- DB : Search'97 (universal server)

8 - Sybase Inc.

Emeryville, California

DB : PowerBuilder

9 - Microstrategy Inc.

Vienna, Virginia

- DB : DSS Server, Warehouse Moniltor DSS Web.

10- Hewlett - Packard

Palo Alto, California

- DB : Open View

11- Tandem Computers

Cupertini, California

12- Bea Systems

Sunnyvale, California

ويقوم هؤلاء المطورون والموردون بتقديم تكنولوجيا متطورة ومتقدمة مصممة لمساعدة المستخدمين فى نشر بياناتهم على شبكة الويب بالإضافة إلى استخدام الويب كوسيلة للتساؤل فى قواعد البيانات الخاصة بهم بكفاءة عالية .

على أنه توجد أيضًا تكنولوجيات متقدمة تساعد المستخدمين الحصول على البيانات من قواعد البيانات واستخدامها بما يعود عليهم بالفائدة . فإلى جانب تواصل وترابط قواعد البيانات مع الإنترنت ، فإن هذه التكنولوجيات تستلهم التغيير فى قواعد البيانات ، كما يلى :

- إنتاج أنواع استخلاص مختلفة وأدوات تحليل جديدة ، كما تمثله تكنولوجيات Data Mining ، OLAP ومنتجات أخرى لتحليل نهاية الواجهة Front - end .
- تطوير قواعد بيانات الجيل التالى المرتبطة بأدوات التأليف والإنشاء على شبكة الويب .

- مساندة أنواع بيانات الوسائل المتعددة ومن ضمنها النصوص .
- القدرة على تداول الأشياء Objects وخاصة قواعد الأعمال المفسرة من قبل .
- الإمداد والتنشيط المتزايد مع تعزيزات أخرى لمستودعات البيانات Data Warehouses ، Data marts ، Data Webs ، Data processing ، Mission - oriented applications . . . إلخ .

(٣) وصلات الإنترنت لإمداد البيانات :

من الملاحظ أن المستخدمين يريدون الوصول إلى البيانات التى يحتاجون إليها بسرعة وبدون تأخير . وبدلاً من مجابهة فيض كبير من المعلومات المطلوب الوصول إليها ، فإن الأطراف الوسيطة من مديرى قواعد البيانات ونظم المعلومات يوفرون المعلومات الملائمة والمناسبة والدقيقة التى يحتاج إليها المستخدمون بدون إسهاب أو إطناب . وبذلك يمكن تحرير وقت المستخدم فى استقراء ما يصل إليه من معلومات ملائمة للقرار والفعل مباشرة بدلاً من تداول الكم الكبير من المعلومات ومحاولة تحليلها للوصول إلى المطلوب من المعلومات مما يستغرق وقتًا طويلاً وجهدًا مضنيا . هذه الوظيفة المناطة بالطرف الوسيط ، أصبح فى الإمكان تدريب المستخدمين الجادين عليها لأدائها بأنفسهم دون وسيط ثالث .

وتضمنت استراتيجيات مطورى نظم إدارة قواعد بيانات النص إضافة إمكانية تواصل وربط قواعد بياناتهم على شبكة الإنترنت مما يعطيها قدرة على الوصول السريع إليها من خلال محركات البحث المتاحة لذلك. وعلى سبيل المثال ، طورت شركة Oracle برمجيات وسيطة Middleware مع تطبيقاتها لدعم ومساندة البحث والوصول إلى المعلومات من خلال الإنترنت . كما طورت شركة Sybase برامج لتنشيط قاعدة بياناتها وتوسعها لكى تساند تكنولوجيا «مستودع البيانات Data Warehouse » الخاص بها على الإنترنت . وقامت شركة Microsoft بتطوير برمجيات تعزيز وترابط مع الإنترنت فى قاعدة بياناتها المتصلة بخادم لغة الاستفسار الهيكلية SQL Server DB بنفس الطريقة التى تضيف بها قدرات الإنترنت لكل تطبيقاتها ومنتجاتها المرتبطة بنظام التشغيل الخاص بها . وقد أدى كل ذلك إلى تحسين الطريقة التى ينشر بها الناس بياناتهم على الإنترنت مستخدمين «خادم لغة الاستفسار الهيكلية SQL Server» ، وخادم معلومات الإنترنت وغير ذلك من المنتجات .

وقد أصبح فى الإمكان استخدام أداة Internet Wizard فى خادم لغة الاستفسار الهيكلية لنشر التساؤل كصفحة من صفحات الإنترنت بلغة HTML .

وتقوم استراتيجية شركة Sybase على الترابط والتواصل مع الإنترنت ، حيث أن «مستودع بياناتها Data warehouse ذا السمة التفاعلية مصمم أيضًا لإنتاج تساؤلات سريعة وتفاعلية باستخدام منتج Sybase's IQ . وقد جعلت الشركة خاصية الارتباط مع الإنترنت التوجه الرئيسى لكل إصداراتها الحديثة الخاصة بقاعدة بياناتها System II التى تساعد تطبيقات المكتبة أو مركز المعلومات مع الإنترنت مما يتطلب توافر أدوات توفر للمستخدمين إجابات فى بيئة تصفح على الإنترنت ، وتقدم لهم برمجيات وسيطة ملائمة تساند وصل قاعدة البيانات مع الإنترنت .

وتضع شركة Oracle الأسس الحاكمة للبرمجيات الوسيطة والمحركات والأدوات التى تسمح لمستخدمى نظام إدارة قاعدة بياناتها الوصول إلى البيانات التى يحتاجون إليها والمحملة على إصدارتها المختلفة . وتوجد فى نطاق استراتيجية الإنترنت مجموعة تكنولوجيات متقدمة لتأكيد التصرفات التى تحدث فى الوقت الحقيقى على الويب Web والتى تدعمها برمجيات Web Server 3.0 التى ترتبط بعائلة إصدارات الويب لقاعدة بيانات Oracle التى تستطيع تهيئة لغة HTML فى الوقت الحقيقى .

(٤) إمتداد وتوسع الاداء : Scalability Performance

تحتاج قدرة وقوة نظام المعلومات إلى مصداقية الأداء وامتداده أو توسعه ، وحتى يمكن تحقيق الحاجة إلى سرعة الأداء وامتداده ، تركز شركات إنتاج قواعد البيانات على خواص التوازى والتطابق والتوافر والمساندة للتصرفات بطريقة أحسن . وعلى سبيل المثال ، توفر قاعدة بيانات Informix تكنولوجيا برمجيات Online XPS ، وتقدم قاعدة بيانات Oracle خادم توازى وتطابق ، وتتضمن قاعدة بيانات Sybase برمجيات MPP ، وتقدم شركة IBM طبعة توازن وتوافق مع قاعدة بيانات DB2 ، وتعمل شركة Microsoft مع شركة NCR وشركات أخرى لتحسين امتداد خادم لغة الاستفسار الهيكلية SQL Server .

وفى الوقت الحالى ، يتجه كثير من المستخدمين إلى استخدام وحدات بيانات صغيرة ومحددة Data marts بدلاً من مستودعات البيانات Data warehouses العملاقة التى تغطى

كل البيانات ، إلا أنه لم يتنبأ هبوطا فى الأداء والامتداد لقواعد البيانات أو مستودعات البيانات العملاقة حتى الآن .

وفى هذا الصدد يمكن ملاحظة أن امتداد وتوسع الأداء يرتبط بالعوامل التالية :

- حاجة كثير من المستخدمين إلى قواعد بيانات متطورة وممتدة تتضمن أساليب مستودعات بيانات Data warehousing ، استخراج البيانات واكتشافها Data mining ، والوسائط المتعددة Multimedia والتطبيقات المساندة من شبكة الويب Web .
- ترابط وتواصل قواعد البيانات مع شبكة الإنترنت أدى إلى تغيير وضعية قواعد البيانات ، حيث يقوم مطوروا وموردوا قواعد البيانات بدعم ومساندة قواعد بياناتهم لكى ترتبط مع الويب ، مما يساعد فى تبادل المعلومات الوثائقية كما يساند التجارة الإلكترونية المعتمدة على قاعدة البيانات المطورة للمنظمة المعينة .
- صارت التطورات المرتبطة باستخراج البيانات واكتشافها ، معالجة التطبيقات على الخط OLAP ، وتحليل نهاية الواجهة Front - end تسمح للمنظمات المختلفة فى الاستجابة لشروط التفاعل والإنتاجية المطلوب التوصل إليها .
- توفر الخادمات العالمية Universal Servers التى تقدم التكنولوجيا المتقدمة المساندة لأنواع البيانات التى من ضمنها النص ، وبذلك تستخدم كل أو معظم قواعد البيانات لإدارة النصوص المتشعبة والضخمة .
- أضافت تكنولوجيا التوجه الشيئى العلاقى Object - orientation relational قدرات مساندة متنوعة لبيانات النص بطريقة سريعة .
- ساعدت أدوات إدارة قاعدة البيانات المديرين على التحكم بطريقة أحسن فى قواعد بيانات النص وقواعد بيانات الوسائل المتعددة مما يساعد المستخدم فى الحصول على المعلومات المحددة .
- مكنت قدرات الامتدادية والأداء لقواعد البيانات فى أن تصبح أسرع وأحسن فيما يختص بالمعالجة والتصرفات المتوازنة لكى تساند مواقع الويب المشتملة على النصوص والوسائل المتعددة .

(٥) استيعاب المتطلبات المستقبلية :

توجد بعض المؤشرات الإيجابية التى توضح أن منتجات قواعد البيانات المرتبطة بالنص قد أدت إلى القدرة على استيعاب متطلبات النص ذات الأشكال والأنماط الجديدة إلى حد ما . وفى الماضى ، لم يكن يتوقع أى شخص الحاجة لكى تقوم قواعد البيانات بتخزين صفحات الويب فى شكل لغة تعليم النص المتشعب Hypertext Markup Lanaguage (HTML) كنوع من أنواع البيانات . إلا أنه فى الوقت الحالى ، أصبح معظم بل كل مطورى نظم إدارة قواعد البيانات العلاقية RDBMS يخططون لتكامل خادمات قاعدة البيانات Database Servers مع الإنترنت والويب .

ويعتمد اختيار الحلول المرتبطة بالنص وقاعدة البيانات على مدى تلبية كل منها لمتطلبات النص وقاعدة البيانات . وصارت كثير من نظم إدارة قواعد البيانات المتاحة حاليًا تخاطب النص والوثائق ككيانات يجب الوصول إليها وإدارتها ؛ ومن هذه النظم ما يلى :

- منتجات بحث النص الكامل Full - text - search .
- نظم إدارة قواعد البيانات العلاقية الممتدة Extended RDBMS .
- نظم إدارة قواعد البيانات العلاقية الشىء ORDBMS .

كما تتوافر أيضًا تطبيقات متخصصة ، مثل : نظم إدارة الوثائق DMS ومذكرات لوتس Louts Notes التى تعمل على تكامل وبناء بحث النص مع تكنولوجيا إدارة قاعدة البيانات .

وتوفر كثير من المنتجات المتاحة حاليًا قدرات بحث النص الكامل . ومن أمثلة هذه المنتجات ، كل من Excalibur ، Fulcrun ، Open Text ، Personal Library ، Software (PLS) ، Verity . . . إلخ . التى تعمل على تداول النص مع إتاحة قدرات محدودة لتكشيف واسترجاع النص المبنى على عناصر بيانات هيكلية . وبصفة عامة ، تقوم هذه المنتجات بعمل بحوث جيدة من خلال النص ، كما أنها تقوم بالقراءة فقط ، وتتطلب استخدام التفاعلات المعينة ، وتتسم بالقدرة على التوزيع المحدود .

وفى هذا الصدد ، يشتمل منتج Fulcrun على لغة استفسار هيكلية SQL ممتدة لإنشاء

الاستفسارات . كما يمكن لبعض المنتجات الأخرى مثل منتج Open Text بحث نص مبنى على خصائصه وهيكليته .

ويواجه المستخدمون على الخط Online بكم متعاظم جداً من المعلومات التى يوفرها مقدموا خدمات المعلومات والإنترنت المتضمنة قدرات تكشيف النص والبحث فى خادمات ويب حتى يمكن مساعدة المستخدمين فى العثور على ما يحتاجونه من معلومات بكفاءة أكبر . وفى هذا الإطار ، توفر منتجات شركات مثل Verity ، Open Text ، PLS وغيرها محركات بحث النص Text Search Engines كأدوات متضمنة تتسم بها . وعلى سبيل المثال . تستخدم PLS محرك بحث خاص بها لخدمة «أمريكا على الخط America on Line (AOL)» ، وتقوم شركة Verity بتصدير دعم برنامج متعهد Agent مع خادم متعهد الموضوع Topic Agent Server الخاص بها . ووفرت شركة Excalibur أداة لا تقتصر على النص فقط ، بل تمتد لتكشيف واسترجاع الأشكال الأخرى من البيانات . وفى هذا التوجه ، تشترك كل من شركة Excalibur وشركة Informix فى تطوير أدوات تكشيف واسترجاع النص والشكل أى الوسيط . كما عملت شركة In Magic على تبنى مدخلا مستقلا لتجميع بعض الخواص الموجهة لقاعدة البيانات مع استرجاع النص فى قاعدة بياناتها DB/Text Works المبنى على برمجيات النوافذ حيث يضيف برنامج كتابة التقرير Report - writes ، ويقدم القدرة على تخزين وعرض وطباعة النص وأكثر من ٣٠ شكلاً .

المراجـــع

1 - "Beyond Full - text Search" Work group computing Report, Vol. 18, No. 12 (December 1995) .

2 - Celko, Joe "Understonding and managing textbases" **DBMS**,Vol. 9, No. 11 (October 1996), pp. 55 - 62 .

3 - Davis, Judith R. "Textbases", **Open Information Systems**, Vol. 11, No. 1 (Jan. 1996), pp. 23 - 38 .

4 - Stodder, David . " The database dozen " **Database Programming & Design**, Vol. 9, No. 13 (Dec. 15, 1996) , pp. 10 - 23.