

الفصل الثالث عشر المستودعات والأرشيفات

يتناول هذا الفصل أساليب اختزان المواد الرقمية في المستودعات وحفظها لمدة طويلة، كما يتناول البروتوكولات التي توفر إمكانية الوصول إلى تلك المواد المختزنة. وعلى اعتبار أن التخزين لمدة طويلة هو أحد العناصر الرئيسية التي تقوم عليها المكتبات الرقمية، فقد يبدو غريباً للوهلة الأولى أننا أرجأنا تناول هذه الموضوعات الهامة إلى هذا الجزء الأخير من الكتاب، إلا أن لدينا مبررات لذلك؛ حيث ينصب تركيزنا في هذا الكتاب على ما هو موجود بالفعل في وقتنا الحاضر. ومع أننا عرضنا لموضوعات بحثية كلما وجدنا ذلك ضرورياً، إلا أن معظم مناقشاتنا كانت منصبة على النظم المستخدمة في المكتبات في وقتنا الحاضر. وفي حقيقة الأمر، ليس هناك اتفاق كامل في وجهات النظر حول القضايا التي سنعرضها في هذا الفصل حتى الآن، فباستثناء خادم الويب، هناك شبه إجماع حول مستودعات المكتبات الرقمية، كما أن الأرشفة الرقمية أو الحفظ الرقمي لا يزال مجالاً جديداً. وإن كانت قد ظهرت بوادر على بدء فهم الاحتياجات المرتبطة بهذه العملية، فإن أساليب العمل لم تتبلور بعد.

المستودعات :

المستودع هو نظام حاسوبي آلي تتمثل وظيفته الأولى في اختزان المواد الرقمية لاستخدامها في مكتبة ما. وعلى هذا الأساس فالمستودعات هي بمثابة أرفف الكتب في المكتبة الرقمية. وهذه المستودعات قد تكون ضخمة

وقد تكون صغيرة، تختزن ملايين من الكائنات الرقمية أو تقتصر على كائن رقمي واحد فقط. وفي بعض الحالات يمكن اعتبار أية وسيلة متنقلة تشتمل على قليل من الكائنات الرقمية مستودعاً، إلا أن معظم المستودعات تتألف من نظم حاسوبية آلية ضخمة تختزن المعلومات في نظام ملفات أو في قواعد بيانات وتتيحها إلى العالم عبر واجهة محددة ومعروفة.

خوادم الويب :

تُعد خوادم الويب النمط الأكثر شيوعاً للمستودعات في وقتنا الحاضر. وهناك شركات عديدة توفر هذه الخوادم الجيدة. وتكمن الفروق الرئيسية بين منتجات هذه الشركات في البرامج المصاحبة لهذه الخوادم مثل البريد الإلكتروني، وبرامج التكشيف، وأنظمة الأمن، وآليات التسديد المالي الإلكتروني، وغيرها من الخدمات الأخرى للشبكة.

أما فيما يتعلق بالمكتبات الرقمية، فإن خوادم الويب تؤدي مهام مقبولة بتكاليف منخفضة، وقد ساهمت هذه الخصائص في تقبل هذه الأجهزة على نطاق واسع وظهور مستوى يكاد يشكل قاعدة للتشغيل المتداخل. وتدين الويب بالفضل في جانب كبير من نجاحها إلى بساطتها وسهولتها، وما الحاسبات الخادمة إلا جزء من هذا النجاح، ومع ذلك يعد جانب من هذه البساطة التي تميز هذه الأجهزة غير ملائمة أو غير مريحة للقائمين على تنفيذ المكتبات الرقمية. فالحاسبات الخادمة لا تدعم سوى نموذج كائني واحد فقط: وهو نظام الملف الهرمي الذي تنظم فيه المعلومات في ملفات منفصلة، ومعالجتها لهذه الملفات تعد معالجة غير كافية؛ حيث يتم تلقي كل

رسائلها ومعالجتها، وأخيراً نسيانها.

اللوحة رقم (١٣-١)

خوادم الويب

خادم الويب هو برنامج آلي تتمثل مهامه في اختزان الملفات والاستجابة للطلبات الواردة من بروتوكول نقل النص الفائق (HTTP) والبروتوكولات الأخرى المصاحبة له. ويتم تشغيل خادم الويب على حاسب آلي مرتبط بالإنترنت، ويمكن أن يكون هذا الحاسب مخصصاً لهذا الغرض فقط، أو أن يكون جهازاً مشتركاً يقوم بإجراء تطبيقات أخرى، أو أن يكون حاسباً شخصياً يتيح موقعاً صغيراً على الويب.

وفي قلب خادم الويب هناك عملية تسمى " httpd " [الحارس]^(١)، مع ملاحظة أن الحرف d الأخير يشير لكلمة daemon التي تعني شيطان. والحارس هو برنامج يعمل بشكل دائم، إلا أنه يبقى ساكناً معظم الوقت حتى تصله رسالة لمعالجتها. ويعمل بروتوكول نقل النص الفائق على قمة بروتوكول التحكم بالإرسال عبر الإنترنت TCP. ويوفر بروتوكول التحكم بالإرسال منافذ (عناوين) عديدة لكل حاسب، بحيث يرتبط الخادم بأحد هذه المنافذ، وهو المنفذ (رقم ٨٠) في العادة، وإن كان يمكن بالطبع تخصيص منافذ أخرى. وعندما تصل رسالة إلى هذا المنفذ، يقوم بتمريرها إلى الحارس، ومن ثم يبدأ الحارس في التعامل مع هذه الرسالة لتوزيعها، ويتقرب في الوقت نفسه وصول رسائل أخرى. وبهذه الطريقة يمكن تداول

(١) اقتراح من المترجمين.
المكتبات الرقمية

رسائل عديدة في الوقت نفسه من دون إشغال الحارس بتفاصيل معالجتها.

ويتحكم بروتوكول نقل النص الفائق بشكل جيد في المعالجة الفعلية التي يجريها خادم الويب للويب. ولم تكن أوائل خوادم الويب تقوم بأكثر من تنفيذ أمر "توصّل إلى أو احصل على get"، وهذا الأمر يعني تلقي رسالة تحتوي على محدد موحد لموقع المصدر URL من أحد الحاسبات العميلة، ويقوم هذا المحدد الموحد بتحديد ملف مخزن في الخادم، ومن ثم يسترجع الخادم الملف ويعيده إلى العميل. وبهذا تنتهي علاقة بروتوكول نقل النص الفائق بتلك الرسالة.

ومع ازدياد إمكانات بروتوكول نقل النص الفائق، وازدياد أعداد مواقع الويب أصبحت خوادم الويب أكثر تعقيداً. فعلاوة على تنفيذ مجموعة كاملة من أوامر بروتوكول نقل النص الفائق؛ أصبحت الأنظمة الخادمة تدعم نصوص واجهة بوابة المرور الموحدة CGI، والامتدادات الأخرى. ومن المتطلبات التي ما زال يتعين على الخوادم (ومتصفحات الويب كذلك) تلبيةها مواصلة دعم الإصدارات القديمة من بروتوكول نقل النص الفائق، بحيث يجب أن تكون هذه الخوادم جاهزة لتلقي أي رسائل بأي إصدار من هذا البروتوكول ومعالجتها على نحو سليم. وعلاوة على ذلك، تواصل الويب إضافة إمكانات الأمن أو الحماية. كما أن الإصدار (رقم 1,1) من هذا البروتوكول يحتوي كذلك على وصلات دائمة تسمح بمعالجة أوامر بروتوكول نقل النص الفائق عبر وصلة واحدة لبروتوكول التحكم بالإرسال.

وتحتاج المواقع الكبيرة على الويب إلى أكثر من حاسب نظراً لأنشطتها

الكثيرة. وهناك وسائل عديدة تستخدم للمشاركة في تحمل العبء، منها طريقة مباشرة تتمثل في إعادة نسخ البيانات على عدة أنظمة خادمة متطابقة، وتعد هذه الطريقة ملائمة ومريحة عندما يكون عدد التساؤلات كبيراً، وحجم البيانات متوسطاً. وهناك أسلوب آخر يسمى "خدمة اسم النطاق الدوارة DNS round robin" يستخدم لموازنة الحمل من خلال توسعة نظام اسم النطاق بحيث يسمح له أن يشير إلى مجموعة من الحاسبات ذات عناوين مختلفة لبروتوكول الإنترنت، وعلى سبيل المثال، يشير اسم النطاق "www.cnn.com" إلى مجموعة حاسبات، يوجد بكل واحد منها نسخة من موقع الويب الخاص بشبكة سي إن إن CNN، وعندما يدخل المستخدم إلى هذا الموقع يقوم نظام اسم النطاق باختيار أحد الحاسبات لخدمة الاستفسار، إلا أن استنساخ الموقع لا يكون مفيداً عندما يكون حجم البيانات ضخماً، أو إذا كانت البيانات سريعة التغير. وتعد خدمات بحث الويب مثلاً جيداً على ذلك؛ فبعض نظم البحث في الويب تستخدم حاسبات منفصلة لتنفيذ عملية البحث أو إجرائها، ثم تجميع الصفحات التي سترجع أو تقدم إلى المستخدم، هذا فضلاً عن استخدامها لأغراض عرض الإعلانات.

المستودعات المتطورة :

على الرغم من أن خوادم الويب تُستخدم على نطاق واسع، إلا أن هناك أنواعاً أخرى من أنظمة التخزين تُستخدم كمستودعات في المكتبات الرقمية. ففي معالجة بيانات قطاع الأعمال، تعد قواعد البيانات الارتباطية relational databases هي الطريقة المقننة في إدارة الكميات الكبيرة من البيانات. وتبنى قواعد البيانات الارتباطية على أساس نموذج كائني يتألف من قوائم البيانات

من جهة، وطبيعة العلاقات الموجودة بينها من جهة أخرى. وهذه العلاقات تسمح بجمع بيانات من قوائم مختلفة بطرق عديدة؛ فالقوائم وحقول البيانات الموجودة داخل قاعدة البيانات الارتباطية يتم تعريفها من خلال خطة وقاموس للبيانات. وتعد قواعد البيانات الارتباطية جيدة في إدارة كميات كبيرة من البيانات عندما تتم في سياق بنية محددة أو معرفة بشكل جيد. وعليه فإن العديد من دور النشر الكبرى تعتمد على قواعد البيانات الارتباطية في إتاحة موادها، مدعومة بخوادم ويب تقوم بتوفير واجهات تربط بين المجموعات والمستفيدين.

أما بالنسبة لفهارس المكتبات الرقمية وكشافتها فإنها عادة ما تتاح على نظم البحث التجارية، بحيث يكون لهذه النظم مجموعة من الكشافات التي تحيل إلى الكائنات الرقمية. وفعالياً يكون لهذه النظم نموذج مرن ومتقدم لتكشيف المعلومات، ونموذج آخر بدائي للمحتوى الفعلي. وكانت بدايات العديد من هذه النظم على شكل نظم النصوص الكاملة، وكان أقوى ما تقوم به هو توفير إمكانية لاسترجاع أحجام كبيرة من النصوص. وقد أضافت بعض النظم الأخرى إمكانات التغذية الراجعة ذات الصلة بالموضوع، بالإضافة إلى عمليات البحث الحقلي، وإمكانات أخرى، متطلعة بذلك إلى زيادة قدراتها الوظيفية وبالتالي زيادة مبيعاتها.

إن قواعد البيانات الارتباطية ونظم البحث التجارية توفر أدوات جيدة لتحميل البيانات، وتوثيقها، ومعالجتها، وحمايتها. كما تتسم عمليات إدارة الوصول بالدقة، وتتوافر كذلك الخدمات ذات الأهمية في التطبيقات التجارية (مسيرات مراجعة الحسابات التجارية). وهناك اتجاه واسع النطاق في المكتبات الرقمية

صناعة نظم قواعد البيانات لإضافة إمكانية بحث النصوص الكاملة، وأن تعمل نظم البحث على توفير بعض أجزاء ومكونات نموذج قاعدة البيانات الارتباطية، وقد تعد هذه المميزات الإضافية مفيدة، إلا أنه حتى الآن لم تتمكن أي شركة من إنتاج نظام يجمع أفضل ما في هذين المنهجين.

وبالرغم من أن بعض المكتبات الرقمية تستخدم قواعد البيانات الارتباطية بنجاح، فإن النموذج الارتباطي للبيانات ليس مرناً بما فيه الكفاية بالنسبة للنماذج الكائنية القوية الناشئة. وتجمع المكتبات الرقمية الرائدة فيما بينها على أهمية الحاجة إلى مستودعات أكثر تطوراً. وفيما يلي مجموعة من المتطلبات الضرورية لمثل هذه المستودعات.

إمكانية إخفاء المعلومات : Information hiding

يجب إخفاء التنظيم الداخلي للمستودع عن الحاسبات العميلة، وأن تتوفر إمكانية إعادة تنظيم أي مجموعة من المجموعات أو تغيير وضعها الداخلي، أو نقلها إلى حاسب آخر من دون أن يكون لذلك أية تأثيرات على المجموعات الأخرى.

النماذج الكائنية Object Models :

يجب أن تدعم المستودعات مجموعة مرنة من النماذج الكائنية، مع تقليل القيود المفروضة على البيانات، وما وراء البيانات، والروابط الخارجية، والعلاقات الداخلية. ولا ينبغي أن يؤدي إدخال فئات جديدة من المعلومات إلى إحداث تغييرات جوهرية في بنية الجوانب الأخرى للمكتبة الرقمية.

البروتوكولات والصيغ المفتوحة Open protocols and formats :

يجب أن تتمكن الحاسبات العميلة من الاتصال بالمستودعات عن طريق

بروتوكولات البيانات وأنواعها، صيغ معرفة على نحو واضح، ويجب أن تسمح البنية الفنية للمستودع بإدخال تغييرات جوهرية على البروتوكولات كلما دعا تطويرها إلى ذلك. ولعل هذا أكثر ما ينطبق على عملية إدارة الوصول بصفة خاصة. وأخيراً يجب أن يسمح المستودع بتطبيق مجموعة من السياسات على جميع المستويات.

مدى الموثوقية وكفاءة الأداء Reliability and performance :

يجب أن يكون المستودع قادراً على تخزين كميات كبيرة للغاية من البيانات، وأن يكون موضعاً للثقة الكاملة، وأن يعمل جيداً.

ما وراء البيانات في المستودعات :

تخزن المستودعات البيانات وما وراء البيانات. وما وراء البيانات هذه يمكن تصنيفها في ثلاث فئات عامة هي: ما وراء البيانات الوصفية، وما وراء البيانات البنائية أو الهيكلية، وما وراء البيانات الإدارية، كما قد تكون هناك حاجة إلى إيجاد محددات للتمييز بين عناصر الكائنات الرقمية فضلاً عن محددات وصف الكائنات نفسها. ويتطلب تخزين ما وراء البيانات في المستودعات إجراءات تتسم بالمرونة. وهناك عدة احتمالات لعملية اختزان ما وراء البيانات هذه، تتمثل فيما يلي :

- ما وراء البيانات الوصفية، وتخزن غالباً في فهرس وكشافات تتم إدارتها خارج المستودع. ويمكن الاحتفاظ بهذه الفهارس والكشافات في مستودعات منفصلة، كما يمكن لهذه الفهارس والكشافات أن تمتد تغطيتها لتشمل مواد في عدة مكتبات رقمية مستقلة. أما المحددات فتستخدم في هذه الحالة لربط ما وراء البيانات مع البيانات ذات الصلة المكتبات الرقمية

بها.

• ما وراء البيانات البنائية والإدارية، وتختزن غالباً مع الكائنات الرقمية، وغالباً ما تكون ما وراء البيانات هذه مضمنة فعلياً داخل الكائن نفسه.

• بعض ما وراء البيانات تشير إلى مجموعة من الكائنات، أما ما وراء البيانات الإدارية المستخدمه لأغراض إدارة الوصول فيمكن أن تستخدم أيضاً مع المستودع بكامله أو مع إحدى المجموعات داخل المستودع، كما تستخدم معينات البحث مع العديد من الكائنات.

• يمكن اختزان ما وراء البيانات ككائنات رقمية منفصلة مع إيجاد روابط من الكائنات الرقمية التي تنطبق عليها ما وراء البيانات هذه. وهناك بعض من ما وراء البيانات التي لا تختزن صراحة، ولكن يمكن إيجادها أو إنشاؤها عند الحاجة إليها.

ونظراً لوجود كل هذه الاحتمالات، فإن لكل مكتبة رقمية أفكارها الخاصة بشأن اختيار ما وراء البيانات ووصفها، إلا أن تبادل ما وراء البيانات يبقى قضية القضايا بالنسبة لعملية التشغيل المتداخل. وما نظام العمل أو إطاره المسمى وارويك Warwick، الوارد في اللوحة رقم (١٣-٢) إلا شكل من أشكال التنظيم المقترحة للتعامل مع هذا الموقف المضطرب.

اللوحة رقم (١٣ - ٢)

إطار عمل وارويك The Warwick Framework

يرجع أصل نظام أو إطار عمل وارويك إلى بعض الأفكار التي

المكتبات الرقمية

تمخضت عنها ورشة العمل التي عقدت في رحاب جامعة وارويك في إنجلترا عام ١٩٩٦م. والهدف الأساسي لهذا النظام هو تنظيم ما وراء البيانات. فهناك مجموعة كبيرة من ما وراء البيانات - مثل ما وراء البيانات الوصفية كبيانات الفهرسة المقروءة آلياً MARC، و ما وراء البيانات الخاصة بإدارة الوصول، و ما وراء البيانات البنائية، والمحددات - يمكن أن تطبق على كائن رقمي واحد. وقد اقترح الأعضاء المشاركون في الورشة أن يتم تنظيم ما وراء البيانات في حزم مستقلة، كما هو الحال بالنسبة لحزمة بيانات دبلن كور Dublin Core، والحزمة الخاصة بالبيانات المكانية Geospatial. ولا شك أن لهذا الفصل (لما وراء البيانات) مميزات كثيرة في تيسير عملية التشغيل المتداخل، بمعنى أنه إذا كان كل من الحاسب العميل والمستودع قادرين على معالجة حزم من نوع معين، فإنهما سيكونان قادرين على العمل المتداخل أو المشترك إلى حد ما، حتى إن لم يكونا مشتركين في حزم ما وراء البيانات الأخرى التي يدعمها كل منهما.

ويتبنى كل من كارل لاجوز Carl Iagoze من جامعة كورنيل، ورون دانييل Ron Daniel من مختبر لوس ألاموس الوطني، هذه الفكرة البسيطة، كما طوراً طريقة رائعة لاستكشاف جميع عناصر الكائن الرقمي. وكانت أولى ملاحظتهما أن ثمة غموضاً كثيراً في التمييز بين البيانات و ما وراء البيانات، وضرباً مثلاً لذلك بتساؤلها عن مدى اعتبار قائمة محتويات الكتاب جزءاً من محتواه، أم أنها تعد بمثابة ما وراء بيانات خاصة بمحتوى الكتاب؟ وأبانا أن مثل هذا التمييز ليست له أهمية في نظام وارويك. لأن كل

شيء يتم تقسيمه إلى حزم، ولا يتم التمييز بين البيانات وما وراء البيانات. أما الملاحظة الثانية فتمثلت في أنه ليس من الضروري اختزان كل حزمة على نحو صريح على أنها جزء من كائن رقمي؛ فما وراء البيانات الوصفية غالباً ما تختزن مستقلة وكأنها تسجيلية في أحد الفهارس أو في الكشافات، كما أن الشروط والأحكام التي تطبق على العديد من الكائنات الرقمية تختزن في أفضل حالاتها في تسجيلات ضبط أو تسجيلات تنظيمية منفصلة policy record، وليس بإدراجها ضمن كل كائن رقمي. ويمكن أن يتم تنفيذ هذا الفصل عن طريق السماح بوجود حزم غير مباشرة، حيث تخزن الحزمة في أي مكان مناسب لها، مع وجود إشارة مرجعية لها في المستودع، وهذه الإشارة المرجعية قد تكون بمثابة مؤشر بسيط إلى موقع ما، أو قد يستعان ببرنامج آلي يؤدي تشغيله إلى جلب الحزمة عند الحاجة إليها.

إن الكائنات الرقمية من قبيل النوع البنائي الواحد عادة ما تتألف من مجموعة خاصة من الحزم. وهذا من شأنه إيجاد نموذج كائني يستخدم لأغراض التشغيل المتداخل بين العملاء والمستودعات لهذا النوع من الكائنات الرقمية. ومما تجدر الإشارة إليه أنه بالرغم من أن نظام وارويك لم يطبق على نحو صريح في أي نظام واسع النطاق، فإن الأفكار التي أثارها تبدو واعدة. فتقسيم المعلومات إلى حزم محددة بوضوح يبسط عملية وصف الكائنات الرقمية، كما يوفر المرونة اللازمة لعملية التشغيل المتداخل.

بروتوكولات التشغيل المتداخل :

تحتاج عملية التشغيل المتداخل إلى بروتوكولات يمكن للعملاء

استخدامها لإرسال رسائل إلى المستودعات، ويمكن للمستودعات بدورها استخدامها لإعادة إرسال المعلومات إلى العملاء. وعلى مستوى الأساس، هناك بعض العمليات اللازمة لتخزين المعلومات في مستودع وتوفير إمكانية الوصول إليها. ويتطلب استخدام نظم فعالة أن يكون العميل قادراً على اكتشاف بنية الموضوعات الرقمية. كما أن وجود أنواع مختلفة من الكائنات يتطلب توافر أساليب وصول مختلفة، وقد تتطلب عملية إدارة الوصول توافر ضمانات التثبيت من الهوية أو إجراء مفاوضات بين العميل والمستودع. وعلاوة على ذلك، قد يرغب العملاء في بحث في الكشافات الموجودة داخل المستودع.

وفي الوقت الحاضر، يعد بروتوكول نقل النص الفائق (HTTP)، وهو بروتوكول الدخول إلى الإنترنت، أكثر البروتوكولات استخداماً في المكتبات الرقمية. وهناك بروتوكول آخر يستخدم على نطاق واسع هو بروتوكول تبادل البيانات الببليوجرافية، والمعروف بمعياري Z39.50.

اللوحة رقم (١٣ - ٣)

بروتوكول نقل النص الفائق (HTTP)

تعد رسالة "توصل إلى أو احصل على get" في بروتوكول نقل النص الفائق أمراً من أحد الحاسبات العميلة إلى أحد الحاسبات الخادمة لاسترجاع أية معلومات قام بتعريفها المحدد الموحد لموقع المصدر URL الذي تضمنته الرسالة. وإذا أشار المحدد الموحد لموقع المصدر إلى عملية ما تقوم بإنتاج بيانات، فإن البيانات التي ستنتجها هذه العملية ستكون هي البيانات

المسترجعة.

وتبدأ الاستجابة لأمر "توصل إلى" بترميز الحالة Status code، المكون من ثلاث تمثيلات رقمية، وقد تكون هذه الرموز مألوفة لمستخدمي الويب وكأنها رموز أخطاء، مثال ذلك الرموز ٤٠٤، الذي يسترجع في حالة عدم العثور على المصدر الذي تم تحديد عنوانه عن طريق المحدد الموحد لموقع المصدر. وفي حالة عدم وجود خطأ، تتبع رموز الحالة بمعلومات فنية (تستخدم أساساً لدعم الملقمات والذاكرات الفورية أو مساعدتها caches)، ثم بما وراء بيانات عن جسم الاستجابة response body. وتقوم ما وراء البيانات بإعطاء العميل معلومات عن عدد الوحدات الطولية الخاصة بنوع البيانات، ولغتها، وتشفيرها، والقيمة الاختبارية القصيرة المعروفة بالهاش hash، والتاريخ. ويستخدم العميل ما وراء البيانات هذه لمعالجة جسم الاستجابة، وهو الجزء الأخير من الرسالة الذي هو عادة الملف الذي يشير إليه المحدد الموحد.

وهناك نوعان آخران من رسائل بروتوكول نقل النص الفائق لهما ارتباط وثيق برسالة "توصل إلى"، هما: الرسالة الأمامية head message، وهي تطلب البيانات نفسها التي تطلبها رسالة "توصل إلى"، إلا أن جسم الرسالة نفسه لا يرسل، وهذه الرسالة تُفيد في اختبار روابط النص الفائق من حيث المصادقية، أو إمكانية الوصول، أو التعديلات الجديدة من دون الحاجة إلى نقل ملفات كبيرة. أما الرسالة الأخرى فهي رسالة المركز post message، وهي تستخدم لتوسعة حجم المعلومات التي يرسلها العميل إلى الخادم، وتستخدم رسائل المركز بشكل شائع من قبل العميل لإرسال إحدى المكتبات الرقمية

كتل البيانات، كما هو الحال في صيغ لغة ترميز النصوص الفائقة HTML، التي يمكن معالجتها بعد ذلك عن طريق نص واجهة بوابة العبور الموحدة أو من خلال بعض التطبيقات الأخرى على الخادم.

ويتمثل الاستخدام الأساسي لبروتوكول نقل النص الفائق في استرجاع المعلومات من الخادم، ويمكن استخدامه لتغيير المعلومات على الخادم، كما تستخدم رسالة أخرى تعرف برسالة "ضع put" لاختزان معلومات معينة في أحد المحددات الموحدة؛ أما رسالة "ألغ delete" فتستخدم لإلغاء المعلومات. تجدر الإشارة إلى أن رسالتي "ضع"، و"ألغ" نادراً ما تستخدمان لأن الطريقة المعتادة لإضافة معلومات إلى خادم الويب تكون من خلال برامج منفصلة تعالج المعلومات الموجودة على الخادم، وليس من خلال رسائل بروتوكول نقل النص الفائق المرسلة من الخارج.

لقد أُدخلت تغييرات كثيرة على بروتوكول نقل النص الفائق منذ ابتكاره من أجل إتاحة الفرصة لإيجاد إصدارات مختلفة من هذا البروتوكول في الوقت نفسه، ولتعزيز أدائه على الإنترنت. كما يقوم بروتوكول نقل النص الفائق بالتعرف إلى أن هناك رسائل عديدة تُعالج من قبل كل من الملقمات أو الذاكرات الفورية caches. وتحتوي الإصدارات المتأخرة من هذا البروتوكول على مجموعة متنوعة من البيانات والخدمات التي تساند هذه التداخلات في معالجة الرسائل. وهناك أيضاً أنواع رسائل خاصة مثل رسالة "الخيارات options"، التي تسمح للعميل بطلب معلومات عن خيارات الاتصالات المتاحة، ورسالة "تتبع trace" التي تستخدم لتشخيص المشكلات وفحصها.

وبالرغم من الاتساع الذي شهده بروتوكول نقل النص الفائق مع مرور الأيام، فإنه لا يزال بروتوكولاً بسيطاً. وقد نجح مصمموه في مقاومة الضغوط الهادفة إلى إضافة مزيد من الخصائص إليه، كما تمكنوا في الوقت نفسه من إدخال تعديلات عليه لتحسين أدائه. وقد لا يتفق شخصان في الرأي حول ما يجب أن يقدم هذا البروتوكول، إلا أن هذا البروتوكول يبقى بلا شك أحد أكبر النجاحات في عالم الإنترنت.

البرمجة الموجهة للكائنات Object-oriented programming والكائنات الموزعة :

يهدف أحد اتجاهات البحث إلى تطوير أبسط بروتوكولات المستودعات الممكنة لدعم الوظائف الضرورية. وإذا كان بروتوكول المستودع بسيطاً، فإن المعلومات عن أنواع الكائنات المعقدة يجب أن تكون مضمنة في الكائنات الرقمية، وهذا الأسلوب يعرف اختصاراً بسودا SODA، أي "الكائن الذكي والأرشفيات الصماء smart object, dumb archive".

وهناك العديد من المشروعات المتقدمة تقوم على تطوير بنيات فنية تستخدم المفهوم الآلي للكائنات الموزعة. وفي هذا السياق نجد أن كلمة "كائن object" لها معنى فني دقيق يختلف عما تعنيه الكلمة في سياق مصطلحي "الكائن الرقمي" و "الكائن المكتبي" المستخدمان في أماكن أخرى من هذا الكتاب. ففي مجال التحسيب الحديث، يعد الكائن جزءاً مستقلاً من شيفرة الحاسب والبيانات المصاحبة لها، والتي يمكن استخدامها وإعادة استخدامها في سياقات عديدة. وعادة ما تكون المعلومات والآليات بداخل الكائن مغلقة encapsulated، مما يجعل جميع التفاصيل الداخلية تبدو مختفية.

وكل ما يعرفه العالم الخارجي عن فئة من الكائنات هو مجرد واجهة عامة تتكون من أساليب (أي العمليات التي تجري على الكائن) وبيانات فورية. وقد يتفاوت تأثير أسلوب معين من فئة إلى أخرى. فعلى سبيل المثال، في المكتبة الرقمية قد يكون لطريقة الإحضار أو الترجمة render معاني أو تفسيرات مختلفة لفئات مختلفة من الكائنات.

وبعد عقود من التطور، أصبحت لغات البرمجة الموجهة للكائنات، مثل لغتي : سي بلص بلص ++C، وجافا، مقبولة باعتبارها أفضل الوسائل المنتجة لبناء النظم الآلية. ولقد كانت القوة الدافعة وراء البرمجة الموجهة للكائنات هي تعقد عمليات التحسيب. يضاف إلى ذلك أن البرمجة الموجهة للكائنات تسمح بتطوير المكونات واختبارها على نحو منفصل، بحيث لا يتطلب تنقيحها لإصدارات متتابعة من النظام. وتعد شركة مايكروسوفت أحد المستخدمين الهامين للبرمجة الموجهة للكائنات. وهناك إصدارات عديدة من بيئة البرمجة الموجهة للكائنات أطلقتها شركة مايكروسوفت تعرف بـ OLE، و COM، و DCOM، و Active X، وكلها أشكال متنوعة للمفاهيم الأساسية نفسها.

إن الكائنات الموزعة تدعو لتعميم فكرة الكائنات على البيئة الشبكية. والفكرة الأساسية تتمثل هنا في أن الكائن الذي يجري تنفيذه على أحد الحاسبات يجب أن يكون قادراً على التفاعل مع كائن آخر بحيث يُنفذ على حاسب آخر عبر واجهته المنشورة، والمعروفة من حيث أساليب التعامل والبيانات الفورية التي تقدمها. وقد طورت شركات البرمجيات الرائدة (باستثناء شركة مايكروسوفت) معياراً للكائنات الموزعة يعرف باسم كوربا

CORBA (١). ويوفر هذا المعيار لمطوري أنظمة نظم الحاسبات الموزعة وسائل الراحة البرمجية نفسها التي تقدمها البرمجة الموجهة للكائنات داخل حاسب آلي واحد.

وتتمثل الفكرة الأساسية لهذا المعيار في وجود وسيط طلب الكائن (Object Request Broker (ORB)). وعند إضافة هذا الوسيط إلى أحد برامج التطبيقات فإنه يقيم علاقات من نوع علاقة العميل / الخادم بين الكائنات. بحيث يتمكن العميل باستخدام الوسيط من إيجاد طريقة على خادم الكائن، وهذه الوسيلة قد تكون موجودة في الجهاز نفسه أو عبر شبكة. ويقوم الوسيط بتلقي الاتصال وإيجاد كائن يمكنه تلبية الطلب، ومن ثم يمرر المؤشرات إلى ذلك الكائن، ويوجد الوسيلة، ثم يسترجع النتائج. ولا يقتضي الأمر أن يكون العميل على علم بمكان وجود الكائن، أو لغته البرمجية، أو نظام تشغيله، أو أي جوانب أخرى من عناصر النظام التي لا تعد جزءاً من واجهة الكائن. وهكذا فإن الوسيط يعمل على توفير التشغيل المتداخل بين تطبيقات موجودة في أجهزة مختلفة تعمل في بيئات موزعة ومختلفة أيضاً.

إخفاء البيانات data hiding :

على الرغم من أن مفهوم الكائنات في سياق مجال التحسيب، يختلف عن مفهوم الكائنات الرقمية في سياق المكتبات، فإن هناك بعض السمات المشتركة بينهما، أما مصطلح "إخفاء البيانات" فهو مستمد من مجال البرمجة الموجهة للكائنات، لكنه ينطبق كذلك تماماً على الكائنات الرقمية في المكتبات. فعندما يدخل أحد العملاء إلى معلومات في أحد المستودعات فإنه

(١) اختصار لـ Computing Object Request Broker Another (المترجمان).

يحتاج إلى معرفة طبيعة الواجهة التي يقدمها المستودع للعالم الخارجي، ولكنه لا يحتاج إلى معرفة كيفية تخزين المعلومات في المستودع. وفي سياق خادم الويب نجد أن الواجهة تتمثل في بروتوكول (بروتوكول نقل النص الفائق HTTP)، وخطة عناوين (المحدد الموحد لموقع المصدر URL)، ومجموعة من الصيغ وأنواع البيانات. أما في سياق مستودعات أخرى، فيمكن التعبير عن ذلك بمصطلحات البرمجة الموجهة للكائنات. وما يعتقد المستخدم أنه كائن رقمي واحد قد يكون مخزناً في المستودع على شكل مجموعة معقدة من الملفات، أو كمجموعة تسجيلات في قوائم، أو كائنات نشطة تنفذ عند الطلب.

إن إخفاء البنية أو التركيب الداخلي، وتوفير جميع إمكانات الوصول من خلال واجهة محددة على نحو جيد، ييسر عملية التشغيل المتداخل. ولا شك أن العملاء قد يستفيدون من عدم احتياجهم إلى معرفة التنظيم الداخلي للمستودعات. فقد يلجأ مستودعان إلى تنظيم معلومات متماثلة بأساليب مختلفة؛ حيث يمكن لأحدهما أن يختزن مسار الصوت وصور الفيلم الرقمي على أنهما كائنان رقميان منفصلان، في حين يختار الآخر اختزانهما على أنهما كائن واحد. ولذا يجب أن يكون برنامج العميل قادراً على إرسال طلب لبدء القراءة أو التشغيل غير مكترث بهذه الاختلافات الداخلية. كذلك تستفيد المستودعات من ذلك لأن التنظيم الداخلي يبقى شأناً داخلياً محضاً. وما يبدو للمستخدم على أنه كائن رقمي واحد، قد يكون في حقيقة الأمر صفحة في صيغة لغة ترميز النص الفائق مع صور مرتبطة بها مع تطبيقات

بلغة جافا. ومع طريقة إخفاء المعلومات، يمكن نقل الصور إلى موقع آخر أو التحويل إلى نسخة أو إصدار جديدة بلغة جافا من دون أن يدري العملاء عن ذلك.

وتوفر كل من خدمة جستور JSTOR، وبرنامج المكتبة الرقمية الوطنية في مكتبة الكونجرس، صوراً منمنمة كإصدارات مصغرة scaled-down versions لصور أكبر. وقد قررت مكتبة الكونجرس إنتاج الصور المنمنمة مقدماً وتخزينها على أنها بيانات مستقلة. أما خدمة جستور فإنها لا تختزن الإصدارات المصغرة، وإنما تقوم بدلاً من ذلك باستخراجها عند الطلب من الأشكال المخزنة للصور الأكبر، وكلا الأسلوبين مقبولان. وهذه القرارات الداخلية يجب أن ألا تظهر على واجهة المستخدم، كما يمكن أن يتم تغييرها في أوقات لاحقة؛ فالنظم الداخلية تحتاج إلى معرفة أن المستودع قادر على تزويدها بصورة منمنمة لا إلى معرفة كيفية استخدامها.

النظم الموروثة Legacy system :

اتبعت المحاولات النمطية للتشغيل المتداخل مساراً يعتمد على معايير متفق عليها. وهذه السياسة تتمثل في إقناع جميع الهيئات بالاتفاق على مجموعة من المعايير الفنية والتنظيمية. أما بالنسبة للنظم المعقدة كالمكتبات الرقمية، فالمهمة شاقة، وهناك حاجة للمعايير لضبط العمل في مجالات كثيرة كبناء الشبكات، وأنواع البيانات، وأساليب التعريف، وسبل الأمن، وأساليب البحث والاسترجاع، والكشف عن الأخطاء، وتبادل دفع رسوم التكاليف، وكل معيار من المعايير الموجهة لهذه العمليات يتكون من أجزاء عديدة، تتناول

الإجراءات المرتبطة بعمليات الصياغة اللفظية والدلالية، وإجراءات تفحص الخطأ، وإعداد التوسعات، وغير ذلك. وإذا تم الاتفاق على مجموعة كاملة من المعايير، وإذا طبقت الهيئات هذه المعايير تطبيقاً كاملاً، فإننا قد نتمكن من تحقيق مستوى رائع للتشغيل المتداخل.

وإذا تأملنا واقع التطبيق العملي، نجد أن السرعة التي تسير بها خطى عملية المعايرة أبطأ من سرعة التطور التقني، إذا لم تتمكن هيئة واحدة من تطبيق المعايير على نحو كامل في أنظمتها، بل حتى قبل أن نبدأ هذا التطبيق الكامل تسارع الهيئات إلى تغيير أنظمتها للاستفادة من فرص جديدة أو لتفادي مشكلات جديدة. والسؤال الذي يطرح نفسه هنا فيما يتعلق بالتشغيل المتداخل هو: كيف تتمكن نظم من أجيال مختلفة أن تعمل بعضها مع بعض؟، وإن كان يطلق على النظم القديمة في بعض الأحيان اسم "النظم الموروثة" - استخفافاً بها، فإن هناك كثيراً من النظم القديمة لا تزال تقوم بأعمال جيدة، ولذا يتعين دائماً أن تستوعب الخطط المستقبلية النظم والالتزامات الحالية.

الحفظ التاريخي أو الأرشيفي Archiving :

الأرشيفات هي المادة الخام للمادة للتاريخ، وتتجسد مهمة إدارة الأرشيف الوطني والوثائق الجارية في الولايات المتحدة الأمريكية في حفظ الوثائق "طوال حياة الدولة". وعلى أقل تقدير يجب أن تكون الأرشيفات مجهزة للقيام بعمليات حفظ المعلومات المكتبية لفترات أطول من أي نظام آلي آخر موجود اليوم، وأطول أيضاً من أية وسيلة إلكترونية أو ممغنطة تمت

تجربتها. ولا شك أن عملية الحفظ الأرشيفي الرقمي مهمة ليست باليسيرة. وقد يكون طرح القضايا أسهل من علاجها، وهناك عدة أمثلة جيدة على ذلك. ويعد التقرير الصادر عن لجنة العمل الخاصة بالحفظ الأرشيفي للمعلومات الرقمية Task force on Archiving of digital information، الوارد في اللوحة رقم (١٣ - ٥)، هو الأساس الذي يقوم عليه العمل الحديث في مجال الحفظ الأرشيفي الرقمي.

وتتميز عملية الحفظ الأرشيفي التقليدي بين مصطلحي الحماية Conservation الذي يعنى بالحفاظ المادي على المواد، والحفظ أو الصيانة Preservation، الذي يسعى للمحافظة على المحتوى حتى في حالة تلف المادة الأصلية أو دمارها. وهناك أسلوبان مناظران لهما في سياق الحفظ أو الأرشفة الرقمية، هما: التحديث أو التنشيط refreshing الذي يهدف إلى المحافظة على التسلسل الدقيق للبتات، والنقل migration الذي يسعى إلى المحافظة على المحتوى بمستواه الدلالي أو اللفظي، ولكن من دون المحافظة على التسلسل الدقيق للبتات. وأول من أقر هذا التمايز هي لجنة العمل الخاصة بالحفظ الأرشيفي للمعلومات الرقمية، التي أوصت باتخاذ "النقل" أسلوباً أساسياً في عملية الأرشفة الرقمية.

ويتطلب التنشيط والنقل جهوداً متواصلة؛ فوثائق الأعمال تحفظ لفترات زمنية طويلة لأن وراءها فريقاً من الأفراد يأخذون أجوراً للمحافظة عليها وصيانتها، وهذا هو عملهم. وهؤلاء الأفراد يعتنون بالقضايا المرتبطة بهذه العملية مثل الأمن، والنسخ الاحتياطي، وتوفير البيانات على المدى الطويل. وقد تنبه الناشرون أيضاً إلى أن معلوماتهم الرقمية تعد إرثاً يمكن أن يدر

عليهم أرباحاً لعقود، إلا أنه، بالنسبة للعديد من المجموعات الرقمية، لا يوجد شخص مسؤول عن المحافظة على المعلومات لفترات تتجاوز مصلحته الأنية. وهناك بيانات قد تبدو اليوم غير ذات فائدة، ولكنها يمكن أن تقدر تقديراً كبيراً في المستقبل البعيد. ولكن الحفظ الأرشيبي لمثل هذه المعلومات يحتل مرتبة دنيا - مع الأسف - في قائمة أولويات كل فرد، وهي أول شيء يستغنى عنه عند تخفيض الميزانية.

اللوحة رقم (١٣-٤)

مشروع إنفوباص في جامعة ستانفورد The Stanford InfoBus

هذا المشروع الذي يعرف بإنفوباص Infobus هو أحد المشروعات التي مولتها مبادرة المكتبات الرقمية وكان يديره هكتور جارسيا - مولينا Hector Garcia -Molina، ثم تيري وينوجراد Terry Winograd، ثم أندرياس بيبك Andrwas Paepake. ويعالج هذا المشروع مشكلة التشغيل المتداخل بين النظم العاملة، من خلال تقبل الأنظمة الموجودة كما هي، دون محاولة إرساء معايير جديدة أو تغيير النظم القديمة. والفكرة الأساسية لهذا المشروع تتمثل في إنشاء ملققات خدمة مكتبية تعرف بكانات كوريا لتمثل الخدمات الإلكترونية المباشرة. وهذه الملققات تقوم بالاتصال بالخدمات الموجودة من خلال أية وسيلة اتصال من الوسائل التي تدعمها، ثم تقوم بتحويل الرسائل إلى واجهات قياسية معرفّة لكوريا، فعلى سبيل المثال، حين يرغب أحد العملاء لديه واجهة بحث يدعمها معيار Z39.50 في البحث في إحدى خدمات البحث المباشرة مثل داياالوج المكتبات الرقمية

DIALOG، فهذا الأمر يتطلب استخدام ملقمين، أحدهما للترجمة بين بروتوكول البحث Z39.50، ونموذج الإنفوباص InfoBus، أما الآخر فللترجمة بين واجهة دايلوج ونموذج إنفوباص. وباستخدام هذين الملقمين يستطيع العميل البحث في دايلوج DIALOG بالرغم من وجود واجهتين مختلفتين.

وربما تكون أكثر أدوات الإنفوباص أهمية هي تلك التي تدعم معيار Z39.50، وقد طورت جامعة ستانفورد ملقماً يمكن الحاسبات التي يدعمها معيار Z39.50 من الاتصال بخدمات البحث التي لا يدعمها هذا المعيار؛ بحيث يمكن للمستخدمين تقديم تساؤلات البحث إلى هذا الملقم عن طريق أي واجهة ربط مصممة للاتصال بخادم يدعمه معيار Z39.50. من ثم يقوم الملقم من خلال الإنفوباص بتمرير طلب البحث إلى أي مصدر من المصادر المتاحة الخاصة بالإنفوباص، حتى وإن كانت هذه المصادر لا تدعم معيار Z39.50، وبعد ذلك يقوم الملقم بتحويل النتائج إلى صيغة يمكن فهمها من قبل الحاسبات العميلة التي يدعمها معيار Z39.50. وقام الباحثون في جامعة ميتشجان بعمل مواز لذلك؛ حيث قاموا بعمل ملقم آخر يساعد على إمكانية الوصول إلى جميع الحاسبات الخادمة التي يدعمها المعيار Z39.50 عن طريق الإنفوباص. وعلاوة على ذلك تبني مشروع جامعة ستانفورد إنشاء ملقومات للتعامل مع بروتوكول نقل النص الفائق، وكذلك للخدمات الأخرى على الويب (بما في ذلك كون تكست Con TEXT، وأداة تلخيص الوثائق في لغة أوراكل Oracle's Document-Summarization Tool).

اللوحة رقم (١٣ - ٥)

لجنة العمل الخاصة بالحفظ الأرشيفي للمعلومات الرقمية

شُكِّلت لجنة العمل الخاصة بالحفظ الأرشيفي للمعلومات الرقمية في أواخر عام ١٩٩٤م من قبل كل من لجنة الحفظ والوصول، ومجموعة المكتبات الرقمية، بهدف دراسة المشكلات التي تواجه الحفظ الرقمي. وترأس هذه اللجنة جون جاريت John Garrett، الذي عمل فيما بعد في اتحاد مبادرات البحوث الوطنية المعروف بسي إن آر آي CNRI، ودونالد وترز Donald Waters من جامعة ييل Yale University. وكان التقرير الصادر عن هذه اللجنة عام ١٩٩٥م هو أول نظرة شاملة على قضايا الحفظ الرقمي من حيث الجوانب القانونية، والاقتصادية، والتنظيمية، والفنية. وقد ألقى هذا التقرير الضوء على مخاطر التعطل التقني، مع ضرب أمثلة مثيرة على ذلك. فقد لاحظ التقرير - على سبيل المثال - أنه لم تعد هناك برامج آلية تستطيع مثلاً تحليل البيانات التي جمعها مشروع مخزون الموارد الوطنية واستغلال الأراضي في نيويورك the New York Land use and resources عام ١٩٦٠م. ودرست لجنة العمل هذه إمكانية بناء نظام أرشيف وطني يتخذ أساساً للحفظ طويل الأمد. وبالرغم من أن اللجنة أقرت بأهمية - بل بضرورة - هذا النظام، إلا أنها أدركت أيضاً أنه لن يكون حلاً شاملاً لجميع المشكلات، فالمسؤولية الأساسية لحفظ المعلومات تقع على عاتق الأفراد والمؤسسات التي تديرها على حد سواء، وهؤلاء يجب أن يعتبروا أنفسهم جزءاً من النظام الأكبر.

وأكد التقرير على أن الحفظ الأرشيفي للمعلومات هو عملية أكبر من

مجرد نقل المعلومات من وسيلة تخزين متقدمة إلى أخرى، بل يجب أيضاً حفظ وسائل تحويل المعلومات نفسها وترجمتها ومعالجتها. ولهذا السبب من المحتمل أن يصبح نقل المعلومات من شكل إلى آخر ومن نظام آلي إلى آخر هو الأسلوب السائد في عملية الحفظ الرقمي.

ومن اللافت للنظر عناية التقرير بدراسة الجوانب القانونية لعملية الحفظ، حيث إن جانباً كبيراً من المعلومات التي ينبغي حفظها مملوك لهيئات تعدها ملكية فردية سواء كان ذلك لدوافع مالية أو لأية أسباب أخرى. وبطبيعة الحال تتردد هذه الهيئات في توفير نسخ من معلوماتها لحفظها في الأرشيفات. ومن ناحية أخرى، يجب أن تؤكد الأرشيفات على حقها في الحصول على المعلومات للصالح العام، وخاصة المعلومات المعرضة للضياع، الخاصة بإحدى الشركات التي لم يعد لها وجود فعلي في السوق.

إن هذا التقرير يعد دراسة كاملة أعدت في هذا المجال. بالرغم من أن الحسابات المالية الواردة فيه قد أضحت قديمة، هذا فضلاً عن أن التقرير لم يتطرق جدياً إلى تحليل الخيارات والفرص المتاحة.

التخزين :

كان طول عمر المنتج المادي a physical artifact في الماضي يعتمد على طول عمر المادة أو المواد التي صنع منها، وسواءً أكان المنتج مجرد سجلات بسيطة صادرة عن الكنائس والحكومات، أو كنوزاً مثل حجر رشيد، ومطويات [مخطوطات] البحر الميت، وسفر يوم القيامة، وأناجيل جوتنبرج، فإن ما تبقى من هذه المنتجات هي تلك التي صنعت من مادة لا

تتلف بسهولة، وعلى الأخص الورق فائق الجودة.

لا يمكن لأحد أن يضمن بقاء أي من الوسائط الرقمية الحالية لفترة طويلة. بل إن العمر الافتراضي لبعضها، كالشريط المغنط، يكون قصيراً للغاية، مع أن هناك بعض الوسائط الأخرى، كالأقراص المدمجة، تكون أكثر تحملاً، إلا أنه لا يمكن لأحد التنبؤ بالعمر الافتراضي الأقصى لها. وعلى هذا فإن جميع المعلومات الرقمية ستكون عرضة للضياع في غضون عقود قليلة إن لم يتداركها أحد بالاهتمام. وتصف اللوحة رقم (٦-١٣) بعض الأساليب المستخدمة لتخزين المعلومات الرقمية في الوقت الحالي، مع ملاحظة أن الاهتمام من نصب على خفض تكلفة الأجهزة والمعدات وعلى زيادة سرعة استرجاع المعلومات، وليس على إطالة عمر المنتج.

اللوحة رقم (٦-١٣)

تخزين المعلومات الرقمية

تستخدم المكتبات الرقمية أعداداً ضخمة من وسائط التخزين؛ فصفاة واحدة من نص مكتوب وفقاً لرموز المعيار الأمريكي لتبادل المعلومات والمعروف بآسكي ASCII، قد تحتوي على بضعة آلاف من الحروف أو التمثيلات characters، في حين أن صفاة واحدة ملونة ممسوحة ضوئياً لا تتعدى مساحتها بوصة مربعة تحتاج أكثر من ميجابايت (مليون بايت)، كما أن ساعة من التسجيل الصوتي المخزن رقمياً على قرص مدمج تحتاج إلى أكثر من ٦٠٠ ميجابايت، بل إن دقيقة واحدة من عرض فيديو قد يحتاج إلى أكثر من جيجابايت من البيانات قبل ضغطها. ولتقليل مساحات التخزين، تُضغط المواد الكبيرة، ويشمل ذلك تقريباً جميع الصور، والتسجيلات

الصوتية، والأفلام. والفكرة الأساسية لضغط الملفات بسيطة مع أن العمليات الرياضية معقدة، فالمعلومات الرقمية تشتمل على فائض من المعلومات الهامشية أو معلومات لا داعي لها redundancy، مثال ذلك أن صورة على صفحة واحدة قد توجد بها مساحات بيضاء غير مستغلة، وعليه فليس من الضروري ترميز كل بكسل فيها على نحو منفصل. وحيث إن الاختلافات بين إطارات الفيديو المتلاحقة طفيفة، فمن الأسهل تسجيل الاختلافات بينها من كل إطار منها على نحو منفصل.

إن الضغط المحكم lossless compression للملفات يقوم بمسح المعلومات المكررة أو الضوضائية على نحو يمكن عكسه تماماً؛ أي يمكن إعادة بناء البيانات الأصلية كما كانت بالضبط، أما الضغط غير المحكم lossy compression فلا يمكن عكسه؛ أي أن التقريب مدعاة إلى فقد بعض المعلومات. وفي بعض التطبيقات يجب أن يكون الضغط محكماً، مثال ذلك في حالة التجارب الفيزيائية، حيث تعد النقطة الواحدة في الصورة ذات أهمية كبيرة، ولذا فإن أي تعديل في الصورة قد يقوّض مصداقية التجربة. وإن كان يمكن في معظم التطبيقات القبول ببعض النواقص. وتعد كل من طريقة JPEG (المستخدمة مع الصور)، و MPEG (التي تُستخدَم للأفلام) من طرق الضغط غير المحكم والتي تجري معايرتهما لتقديم صورة مريحة للغاية للعين البشرية.

إن طرق ضغط الملفات تصغر حجم البيانات إلى حد بعيد، إلا أن الملفات تبقى كبيرة. فحتى بعد الضغط، نجد أن صفحة أحادية اللون monochrome من نص ممسوح ضوئياً يزيد حجمها على ٥٠,٠٠٠ بايت.

كما أن الضغط بطريقة MPEG يصغر حجم الفيلم الرقمي من ٢٠ أو ٣٠ ميجابايت في الثانية إلى ١٠ ميجابايت في الدقيقة. وحيث إن المكتبات الرقمية يمكن أن تخزن ملايين من هذه المواد، فإن سعة التخزين تعد من الأمور الهامة للغاية.

إن وسيلة التخزين المثالية للمكتبات الرقمية يجب أن تسمح بتخزين مساحات واسعة من البيانات وبتكلفة قليلة، وأن تكون سريعة في تخزين المعلومات وقراءتها، وأن تكون محل ثقة وأن تدوم طويلاً.

إن الأقراص الممغنطة الدوارة rotating magnetic desk هي وسيلة التخزين القياسية في النظم الآلية الحديثة. وتتفاوت هذه الأقراص من حيث الحجم بين عدد قليل من مئات الملايين من البايتات إلى وحدات تضم آلاف الجيجابايتات. (الجيجابايت يساوي ١٠٠٠ مليون بايت). كما أن هذه الأقراص سريعة بما يكفي لمعظم تطبيقات المكتبة الرقمية، حيث يمكن قراءة البيانات من الأقراص أسرع من نقلها عبر الشبكات. وعند قراءة البيانات من القرص يكون هناك تأخر بسيط للغاية (حوالي ١٥، ٠ ثانية) وهو الوقت الذي يستغرقه ضبط رؤوس القرص لبدء القراءة؛ ثم تبدأ قراءة البيانات في وحدات كبيرة (حوالي ٢٥ ميجابايت في الثانية في الأوضاع النموذجية). وهذه السمات التي تميز أداء الأقراص تناسب تطبيقات المكتبة الرقمية التي تحتاج لقراءة وحدات ضخمة من البيانات في وقت واحد.

إن انخفاض تكلفة الأقراص يعد إحدى معجزات التقنية؛ فأسعار الأقراص الممغنطة تنخفض تدريجياً أسرع من أسعار أشباه الموصلات، ففي عام ١٩٩٨م كان سعر الأقراص يقدر ببضع مئات من الدولارات لكل

المكتبات الرقمية

جيجابايت. والتقنية تتقدم سريعاً للغاية بحيث يمكن للمكتبات الرقمية أن تخطط بثقة لعشر سنوات قادمة على أساس أن التكلفة لن تتعدى ٥%، ومن المحتمل جداً أن تصبح أقل من ١% من التكلفة الحالية.

أما نقاط ضعف الأقراص فتتمثل في عدم الثقة في الاعتماد عليها، فالبيانات الموجودة على الأقراص يمكن أن تضيع بسهولة، إما بسبب عطل في أحد المكونات المادية للحاسب، أو بسبب احتمال الكتابة عليها عن طريق برنامج آخر. وللوقاية من عملية الضياع هذه، فمن المعتاد نسخ البيانات بانتظام على وسيلة أخرى، عادة ما تكون شريطاً ممغنطاً. ومن المعروف كذلك وجود بعض المساحات غير المستخدمة على الأقراص، لذلك فمن السهل تصحيح الأخطاء البسيطة ألياً. وفي الواقع لا يمكن الاعتماد على الأقراص أو الشرائط الممغنطة للتخزين طويل الأمد. فالبيانات ترمز على فيلم ممغنط رقيق يحفظ على سطح ما، وعاجلاً أم آجلاً يتلف هذا الفيلم، وعليه يمكن القول بأن الأقراص تعد وسيلة ممتازة للعمليات الحالية ولكن ليس للحفظ الأرشيفي.

وأحياناً يتم اختزان مجموعات المكتبات الرقمية الكبيرة في مستودعات ذات طابع هرمي. والمستودع النموذجي يتكون من ثلاثة مستويات: الأقراص الممغنطة، والأقراص البصرية، و الشرائط الممغنطة. فالأقراص الممغنطة تكون دائماً متاحة على الخط المباشر online، بحيث يمكن قراءة المعلومات خلال جزء من الثانية. أما الأقراص البصرية فإنها تعد وسيلة أرخص لتخزين كميات كبيرة من البيانات، إلا أن أسطوانات الأقراص تخزن منفصلة offline، وقبل استخدام القرص البصري، يجب أن يقوم الروبوت المكتبات الرقمية

ROBOT بنقله آلياً من المستودع silo إلى قارئ الأقراص - وهي عملية تتم ببطء. وكذا فإن الشرائط الممغنطة تخزن أيضاً منفصلة، ويتم حملها بواسطة الروبوت.

وبالنسبة للحاسبات التي تستخدم المخازن الهرمية، يعد المخزن الهرمي نظام ملفات مترابطاً واحداً. وقد قل اللجوء إلى اتباع أسلوب نقل البيانات من الأقراص الممغنطة الأسرع والأعلى تكلفة إلى وسائط أبطأ وأرخص. ونتيجة للانخفاض المتواصل في تكلفة الأقراص الممغنطة وسعتها، أصبحت الحاجة إلى إيجاد مستوى تخزين متوسط محل تساؤل. فالأقراص والشرائط الممغنطة تؤدي وظائف جيدة وكلاهما ضروري، إلا أن الحاجة إلى مستوى تخزين متوسط قد تنتفي في المستقبل.

إعادة النسخ والتنشيط Replication and Refreshing :

يعد النسخ أحد الأساليب الأساسية لمعالجة البيانات، فالبيانات الهامة الموجودة كنسخة واحدة فقط على حاسب آلي تكون عرضة للمخاطر، فالمكونات المادية للحاسب قد تتعطل؛ والبيانات قد تضيع بسبب مشكلة في البرامج؛ أو قد يقوم موظف غير مؤهل أو غير أمين بمحو البيانات؛ وقد يحترق المبنى الذي يوجد فيه الحاسب، أو يغمره فيضان، أو تصيبه أية كارثة أخرى. لهذه الأسباب، تلجأ مراكز الحاسب الآلي دورياً إلى إعداد نسخ احتياطية من جميع البيانات وتخزينها في مواقع آمنة. أما الهيئات الأكثر تنظيماً فإنها تزيد على ذلك بتجميع السجلات الهامة دورياً لتخزينها لأمد طويل. ومن أساليب الاستنساخ أيضاً الاحتفاظ بنسخ من السجلات المالية والقانونية على شكل مصغر (أي الميكروفيلم أو الميكروفش)، حيث إن المكتبات الرقمية

الأشكال المصغرة الأرشفية الجيدة تدوم أطول من غيرها.

وحيث إن أعمار جميع الوسائط التي تخزن عليها المعلومات الرقمية قصيرة، فيجب أن تخطط المكتبات الرقمية لتنشيط مجموعاتها دورياً على المنوال نفسه. فكل عدة سنوات يجب نقل البيانات إلى وسائط تخزين جديدة. ومن وجهة النظر المالية، لا يشكل هذا الأمر تحدياً كبيراً، إذ من المتوقع أن يتواصل انخفاض أسعار أجهزة الحاسب الآلي مع زيادة سعتها التخزينية خلال العقود القليلة القادمة. فتكلفة الأجهزة اللازمة لنقل البيانات الحالية ستكون بعد عشر سنوات من الآن نسبة بسيطة من تكلفتها الحالية، ويمكن أيضاً أن تخفف أجهزة الروبوت من ضغوط العمل المطلوبة. وفيما يتعلق بالمكتبات الرقمية، هناك سؤال تنظيمي وهو: هل ستبذل المكتبات والناشرون جهوداً تنظيمية لتنفيذ هذه العمليات؟

حفظ المحتوى عن طريق النقل Preserving Content by

: Migration

حتى لو أخذنا بالافتراض الأكبر المتمثل في تنشيط البتات دورياً من وسيلة إلى أخرى، بحيث يمكن تلبية ضرورة حفظ البيانات الخام أو صيانتها، فإن المشكلات لا تزال في بدايتها. فالمعلومات الرقمية تبقى عديمة الفائدة بدون معرفة ومعالجة قضايا الصيغ والأشكال، والبروتوكولات، وما وراء البيانات. ولا يزال في الإمكان قراءة المخطوطات القديمة، حيث إن اللغات والكتابة تتغير ببطء عبر السنين. وتفسير الوثائق^(١) القديمة يتطلب خبرات كبيرة، وإن كانت هذه الخبرات تقل مع مرور السنين. ولا يستطيع

(١) الأفضل تسميتها تحقيق الوثائق (الترجمان).

الباحثون فك رموز الكتابات القديمة إلا بالمتابعة والإلهام.

إن أنماط تطبيقات الحاسب تتغير على نحو مستمر. فقد تصعب قراءة صيغة ملف عمره ١٠ سنوات، كما لا يوجد في العالم حاسب واحد يمكنه تشغيل برامج لبعض الحاسبات الأخرى التي كانت منتشرة منذ فترة قصيرة مضت. وإن كانت بعض هذه الصيغ بسيطة تماماً. فعلى سبيل المثال، إذا عثر أحد الباحثين الأثريين - مستقبلاً وبالمصادفة - على ملف يتضمن نصاً مكتوباً برموز آسكي ASCII ضمن بيانات أخرى، فمن المحتمل أن يتمكن من قراءة النص وتفسيره، حتى إذا ضاعت جميع المعارف المتعلقة بصيغ رموز آسكي، حيث إن رموز هذه الصيغة بسيطة للغاية. إلا أن هذا الشكل (ASCII) يبقى حالة استثنائية. فهناك صيغ أخرى شديدة التعقيد ومن الصعوبة بمكان أن نصدق أن أحداً يتمكن من فك رموز صيغة MPEG لضغط الملفات من دون أن يتوافر له سجل بالعمليات الرياضية التي بُنيت عليها هذه الصيغة، أو أن شخصاً يمكنه فهم برنامج آلي كبير معتمداً على كود الآلة الخاصة به. لذلك، وعلاوة على تخزين البيانات الخام، يجب أن تحافظ الأرشفة الرقمية على وسائل فهم نوع هذه البيانات وبنيتها والصيغة التي كتب بها، وإذا كان ضرورياً استخدام برنامج آلي لتفسير البيانات فيجب المحافظة على هذا البرنامج وعلى أية وسيلة أخرى يمكنها تنفيذ البرنامج، وإلا فيتعين نقل البيانات إلى أية وسيلة حفظ أخرى.

وعلى المدى القريب، يمكن الاحتفاظ بنظم آلية قديمة لهذا الغرض، إلا أنه يجب الأخذ بعين الاعتبار قصر دورة حياة الحاسبات نفسها. فعاجلاً أو آجلاً،

سيتعطل الحاسب، ولن يتيسر الحصول على قطع الغيار، وأي برنامج يعتمد على هذا الجهاز سيكون عديم القيمة. وعلى هذا يصبح نقل المحتوى ضرورة لازمة.

لقد باتت عملية النقل أسلوباً عملياً قياسيماً في معالجة البيانات منذ عقود. فعلى سبيل المثال، تحتفظ صناديق معاشات التقاعد بسجلات للمعاملات المالية لسنوات طويلة. وفي الولايات المتحدة الأمريكية، تحتفظ إدارة الضمان الاجتماعي بسجلات لضرائب الرواتب المدفوعة عن جميع العمال في كل المهن. وتحفظ هذه السجلات على حاسبات، وتغير النظم الآلية دورياً، حيث يجري إحلال الأجهزة الجديدة محل القديمة، كما يجري تنقيح نظم البرمجيات. وعندما تحدث هذه التغييرات، تنقل البيانات من حاسب لآخر، ومن قاعدة بيانات إلى أخرى. والقاعدة الأساسية التي تستند إليها عملية نقل البيانات، هي أن الصيغ التي توضع فيها البيانات، وبنية البيانات هي التي يمكن تغييرها، أما دلالات المحتوى الأساسي فيجب أن تبقى كما هي.

وهناك أسلوب آخر مقترح لحفظ البيانات، وهو "المشابهة أو المحاكاة Emulation" وتتمثل فكرة المحاكاة في تحديد المواصفات التفصيلية لبيئة استخدام الحاسب اللازم لتنفيذ برنامج معين. ومن ثم يتمكن في أي وقت في المستقبل من بناء بيئة محاكية تعمل مثل البيئة الأصلية تماماً. وفي بعض الحالات الخاصة، تعد المحاكاة اقتراحاً معقولاً؛ فعلى سبيل المثال، يمكن تحديد مواصفات برنامج يستحضر صيغة صورة بسيطة، تشبه صيغة JPEG، ومع ذلك تبقى المحاكاة في جميع الحالات الأخرى مجرد وهم. فحتى بيئات

استخدام الحاسبات البسيطة تعد معقدة للغاية بحيث يصعب تحديد مواصفاتها بدقة. فالدمج الدقيق بين عمليات التركيب والدلالات اللفظية، والقواعد الخاصة أمر يتجاوز الفهم، إلا أن الجوانب الدقيقة الخفية لأي نظام تبقى عنصراً حاسماً للتنفيذ السليم.

الآثار الرقمية القديمة [علم الآثار الرقمي] Digital Archeology :

تمر المجتمعات بفترات عصيبة أثناء فترات الركود والحروب والاضطرابات السياسية، وفي هذه الفترات يصبح نقل المواد الأرشيفية في أدنى سلم الأوليات. ويمكن أن تبقى المنتجات المادية منسية لقرون طويلة في دهاليز الأقبية والمخازن حتى يعاد اكتشافها. أما المعلومات الرقمية فهي أقل تحملاً. وتتناول اللوحة رقم (٧-١٣) كيف صاحب إحدى الفترات العصبية - وهي انهيار ألمانيا الشرقية - فقد أرشيف الدولة. ويطلق على عملية استرجاع المعلومات من مصادر البيانات التالفة غير المكتملة والقديمة، اسم الآثار الرقمية القديمة أو بالأحرى علم الآثار الرقمي.

بناء مكتبات رقمية بالأرشفة الذهنية Archiving in Mind :

بما أن عملية الأرشفة تنطوي على مخاطر كثيرة، فما الذي يمكننا القيام به الآن لتعزيز احتمال تمكن علماء الآثار القديمة الرقميين من تفسير البتات التي اخترعناها؟. هناك بعض الخطوات البسيطة التي من المحتمل أن تشكل فارقاً كبيراً. تتمثل الخطوة الأولى في تخزين المعلومات في صيغ يتبنى استخدامها على نطاق واسع وفي الوقت الحاضر. وهذا سيزيد من فرصة توافر برامج نقل إلى صيغ جديدة في حالة توقف استخدام الصيغ الأولى. فعلى

سبيل المثال فإن صيغتي لغة ترميز النص الفائق HTML، وبي دي إف pdf، تستخدمان على نطاق واسع في عالم الصناعة مما يجعلنا متأكدين من توافر مستخدمين لهما في المستقبل.

ومن الاقتراحات الجديرة بالاهتمام ببناء أرشيف يشتمل على تعريفات الصيغ، ومعايير ما وراء البيانات القياسية والبروتوكولات، والدعائم الأخرى للمكتبات الرقمية. وهذا الأرشيف يجب أن يحفظ على وسيلة حفظ دائمة مثل الورق أو الميكروفيلم، ويجب أيضاً وصف كل شيء بنصوص بسيطة، وفي حالة الحفاظ على الصيغ وأساليب الترميز، فسيمكن فك رموز معظم المعلومات. وقد يعاني علماء الآثار الرقميون في المستقبل من مشقة إيجاد مترجم يمكنه حل الصيغ أو مجموعات التعليقات المهجورة منذ زمن طويل، وإن كان هذا ليس أمراً مستحيلاً. إن صيغ استخدام الحواسيب الحديثة معقدة. ومع أن العالم الأثري الرقمي قد يتمكن من إعادة هندسة البنية الكاملة لأحد الحاسبات القديمة من إنتاج شركة IBM عن طريق استخراج محتويات الذاكرة، فإنه سيقف عاجزاً أمام المواد الأكثر تعقيداً في حالة عدم توافر المواصفات الأساسية للنظام.

وقد يكون "الانتقاء selection"، هو أهم الوسائل التي تستطيع بها المكتبات الرقمية دعم عملية الأرشيف. ولكن ليس مطلوباً ولا ضرورياً الاحتفاظ بكل شيء. فمعظم المعلومات معدة لتكون دورة حياتها قصيرة، ومعظم المعلومات إما أن تكون ثانوية أو عديمة القيمة. كما إن الناشرين يتخذون قراراتهم دائماً بما ينشرون وما يرفضون، وحتى أكبر المكتبات لا

تحصل إلا على نزر يسير من الإنتاج الفكري العالمي. إن المكتبات الرقمية ليست إلا مجموعات معلومات خاضعة لإدارة معينة. وأحد العناصر الأساسية لهذا النظام الإداري هو اتخاذ القرار بشأن ما ينبغي الحصول عليه، وما ينبغي تخزينه، وما ينبغي الحفاظ عليه للمستقبل، وما ينبغي تجاهله واستبعاده.

اللوحة رقم (١٣ - ٧)

علم الآثار الرقمي في ألمانيا

في عام ١٩٩٨م، ظهر مقال في صحيفة نيويورك تايم يتناول التحديات التي يواجهها موظفو الأرشيف في العالم الرقمي في حالة عدم حفظ البيانات على نحو متواصل منذ أول وجود لها.

وفي عام ١٩٨٩م، عندما انهار سور برلين وتوحدت ألمانيا مرة أخرى، كانت السجلات أو الوثائق الرقمية لألمانيا الشرقية في حالة فوضى عارمة، وحصل الأرشيف الاتحادي لألمانيا على مجموعات ضخمة من البطاقات المثقوبة، والأقراص الممغنطة، وأشرطة الحاسب الآلي التي تحتوي على وثائق ألمانيا الشرقية سابقاً. وكانت معظم هذه الوسائط في حالة يرثى لها، وكانت البيانات محفوظة في نماذج غير موثقة، وكانت مراكز الحاسب الآلي التي توجد بها هذه المواد قد أغلقت أو خُصصت على استعجال. ومنذ ذلك الحين، يحاول فريق من الأرشيفيين الألمان إعادة بناء وثائق ألمانيا الشرقية وسجلاتها، وقد أطلقوا على أنفسهم اسم "علماء الآثار الرقميون Digital

"archeologists".

وكانت المشكلة الأولى التي واجهت هؤلاء الموظفين هي استرجاع البيانات من الوسائل المخزنة فيها. أما البيانات المخزنة حتى على أفضل الشرائط الممغنطة فكان عمرها الافتراضي قصيراً، حيث إن الشرائط نفسها كانت في حالة سيئة إلى درجة أن هؤلاء الموظفين لم يتمكنوا من قراءتها سوى مرة واحدة فقط. وفي حالات كثيرة، كانت البيانات مخزنة في نظم روسية Russian system لا تدعمها الحاسبات الآلية الأخرى. وعلى الرغم من أن هؤلاء الموظفين تمكنوا من الحصول على العديد من أجهزة الحاسبات الآلية روسية الصنع، إلا أنهم لم يتمكنوا من قراءة حوالي ٣٠% من البيانات المخزنة.

وعندما نسخت البيانات على وسائل أخرى، كانت المشكلات لا تزال مستعصية الحل، حيث كان هناك جزء كبير من البيانات مضغوطاً بأساليب مبهمة وغير موثقة لغرض توفير المساحة. وقد أثارت إحدى قواعد البيانات الهامة للموظفين الشيوعيين بعض هذه المشكلات. وحيث أن برامج قاعدة البيانات الحاسب الآلية التي حُمِلت عليها هذه البيانات كانت مستنسخة أصلاً من منتجات شركة IBM، فلم يكن استعادة قاعدة البيانات نفسها أمراً صعباً، إلا أن تفسير البيانات في ظل عدم وجود أي توثيق لها كان أمراً شاقاً للغاية. وكانت هناك ميزة واحدة لصالح موظفي الأرشيف، وهي أنه كان بإمكانهم إجراء مقابلات مع بعض الأفراد الذين قاموا بإنشاء قواعد البيانات هذه، واستفادوا من الخبراء من بين هؤلاء الأفراد في تفسير جانب كبير من

المكتبات الرقمية

المعلومات وفي حفظ هذه المعلومات وصيانتها والمحافظة عليها.

وقد لخص مايكل فيتنجيت Michael Wettengel كبير الأرشيفيين الألمان، الموقف بوضوح قائلاً: "إن تقنية الحاسب الآلي وجدت من أجل معالجة المعلومات، لا من أجل اختزانها لآماد بعيدة".