

## الحفظ الرقمي لمواقع الإنترنت

محمود عبد الستار خليفة

ماجستير في المعلومات - جامعة القاهرة

أخصائي معلومات - مكتبة الكونجرس

[Moud@cybrarians.info](mailto:Moud@cybrarians.info)

### تمهيد

أصبحت شبكة الإنترنت أحد أهم مصادر المعلومات الحديثة، حيث يشتمل الإنترنت على كم هائل من المعلومات والتي تأخذ أشكالاً متعددة ومتنوعة كمواقع المؤسسات والخدمات التي تقدمها من تلك المواقع، والدوريات الإلكترونية العلمية، وكذلك قواعد البيانات البيوجرافية وقواعد بيانات النصوص الكاملة. كل هذا جعل الإنترنت أحد أهم مصادر المعلومات الحديثة.

ودائماً ما تأتي التكنولوجيا الحديثة ببعض المشكلات المصاحبة، من أهم تلك المشكلات التي تتعلق بالمعلومات الرقمية بشكل عام هي مشكلة الحفظ، وبالنسبة لمواقع الإنترنت فالمشكلة كبيرة ومعقدة حيث إننا نحتاج إلى حفظ وسيط دائم التغير والتحديث، فمواقع الإنترنت بطبيعتها تتغير بشكل مستمر، وبالتالي فإن ما ينشر فيها من معلومات قد يندثر ولا يمكن الوصول إليه مرة أخرى.

ومن هنا ظهرت تقنية أرشفة مواقع الإنترنت والتي يتم بمقتضاها حفظ نسخة كاملة من مواقع الإنترنت على فترات منتظمة، وإتاحتها للبحث، بحيث يتمكن المستخدم من التعرف على الموقع ومحتوياته كما كان في الماضي، وهذا هو محور الدراسة، والتي تتناول تعريف تقنية أرشفة مواقع الإنترنت، ومراحلها وأجرائها والمشكلات التي تواجهها، ثم تعرض الدراسة أهم تطبيقات هذه التقنية على مستوى العالم، فتتناول أرشيف الإنترنت الدولي وهو أرشيف عالمي لمواقع الإنترنت على مستوى العالم وأول التطبيقات الفعلية لهذه التقنية، ثم تتناول مشروع أرشيف الإنترنت الاستراتيجي، والأرشيف البريطاني.

### ١. أرشفة مواقع الإنترنت Web Archiving

عملية أرشفة مواقع الانترنت هي عملية حفظ مواقع وصفحات الانترنت بهدف الحفاظ عليها من الضياع، والحفاظ على التراث الرقمي لكل دولة، وقد عرف البعض أرشفة الانترنت على إنها عملية تشبه الإيداع القانوني للكتب؛ فمثلما يقوم الناشر بإيداع عدد من النسخ من كل مطبوع ينشره في المكتبة الوطنية بغرض حفظ الإنتاج الفكري الوطني للدولة؛ فإن عملية الأرشفة تسعى إلى الهدف نفسه مع الاختلاف في الشكل والطريقة التي تتم بها.<sup>(١)</sup>

تعريف آخر لعملية الأرشفة؛ هي استراتيجية وعملية تنطوي على مكونات نظرية وتقنية، وتهدف إلى حفظ صفحات الانترنت والمعلومات على المدى الطويل، وذلك لأغراض تاريخية وثقافية وقانونية.<sup>(٢)</sup>

## ١/٢ أهمية أرشفة مواقع الإنترنت

• مواقع الانترنت لها قيمة توثيقية كبيرة، حيث إنها أصبحت مصدرا مهما للبحث في تاريخ وتطور الانترنت كوسيط للمعلومات، حيث إن الانترنت قد تغيرت كثيرا منذ أن ظهرت لأول مرة، وبالتالي فإن انخفاض المواقع القديمة لن يساعد في التعرف على تطور مواقع الانترنت ولغات البرمجة التي استخدمت فيها، ومن هنا تأتي أهمية عملية الأرشفة.

• مواقع الانترنت أصبحت أحد أهم أشكال مصادر المعلومات وبدأت تراحم الأشكال التقليدية لمصادر المعلومات، كما أنها أصبحت قناة مهمة في عملية النشر، ومن هنا أصبح لمواقع الانترنت قيمة معلوماتية كبيرة، وبالتالي هناك ضرورة في الحفاظ على تلك المعلومات خاصة في ظل التغير الدائم في محتوى المواقع.

• بالإضافة إلى القيمة التوثيقية والمعلوماتية لمواقع الانترنت، فإن لها أيضا قيمة ثقافية، خاصة في ظل عمليات رقمنة التراث الوطني للدول، وبالتالي هناك ضرورة في أرشفة تلك المواقع التراثية حتى تظل متاحة في المستقبل.<sup>(٣)</sup>

## ١/٣ مشكلات أرشفة مواقع الإنترنت

تواجه عملية أرشفة مواقع الانترنت أربع مشكلات رئيسية نتناولها فيما يلي<sup>(٤)</sup>:

- النمو المتزايد لحجم الإنترنت

تتسم شبكة الإنترنت بالنمو السريع في اعداد المواقع وفي حجم المعلومات التي تنشر في تلك المواقع، حيث يقدر عدد مواقع الإنترنت المسجلة تحت المستوى الأعلى لأسماء النطاق (Top Level Domains (TLD على مستوى العالم ٤٧١، ٣٨٣، ٦٩، وذلك حتى يونيو ٢٠٠٦<sup>(٥)</sup>، وأدى ذلك إلى صعوبة قيام أى جهة بحفظ ما ينشر على الإنترنت من معلومات.

#### • تطور تكنولوجيا بناء مواقع الإنترنت

تطورت البرمجيات والأدوات التي تستخدم في بناء مواقع الإنترنت، وأدى ذلك إلى تحول المواقع من مجرد صفحات ثابتة المحتوى إلى صفحات يتم انشاؤها عن طريق قواعد بيانات تفاعلية يجب حفظها وأرشفتها.

#### • عدم وضوح مسئولية الأرشفة

لم تتضح بعد ماهية الجهات المنوط بها القيام بمهمة أرشفة مواقع الإنترنت، حيث تتوزع المسئولية بين الأرشيفات الوطنية، والمكتبات الجامعية، بالإضافة إلى بعض المراكز البحثية.

#### • مشكلات قانونية

تعد مشكلة حقوق الملكية الفكرية من أهم المشكلات التي تواجه عملية أرشفة مواقع الإنترنت، حيث يتم في هذه العملية الاحتفاظ بنسخة كاملة من الموقع، وهو ما يعد اختراقاً للملكية الفكرية لأصحاب المواقع، وما يزيد من حدة المشكلة هو عدم تلبية قوانين الملكية الفكرية لاحتياجات عملية الأرشفة، وعلى الرغم من اتجاه بعض الدول لإدخال تعديلات على قانون الملكية الفكرية بما - مثل الدمارك- وذلك حتى يسمح بأرشفة المواقع الوطنية، إلا أنه بشكل عام تمثل حقوق الملكية الفكرية عائقاً أمام عملية الأرشفة، وفيما يتعلق بالجوانب القانونية عند أرشفة المواقع، هناك ثلاثة جوانب يجب وضعها في الاعتبار، وهي<sup>(٦)</sup>:

#### • عملية حفظ المواقع

اعتبرت قوانين حقوق الملكية الفكرية الصادرة في أوروبا والمملكة المتحدة أن صفحات الإنترنت من المواد التي تدخل في إطار التشريعات الخاصة بالملكية الفكرية، وبالتالي فإن أي حفظ أو أرشفة المواقع يعد مخالفاً للقانون، ويلزم الحصول على تصريح من المؤلف أو صاحب الحقوق الفكرية لهذا العمل.

يستثنى من ذلك المواد التي تحمل معلومات شخصية أو مصادر حكومية متاحة للعامة والمصادر التي يكون مسموح نسخها وتداولها.

وفيما يتعلق بالمواد التي يسمح حفظها وأرشفتها بموجب تصريح من الناشر، لابد من توقيع اتفاق مع الناشر

يسمح بحفظ وأرشفة الموقع كاملاً.

### ● إتاحة المواقع المحفوظة للاستخدام العام

تتفق جميع قوانين ومواثيق حقوق الإنسان على حرية الفرد في التعبير عن رأيه وحرية الحصول على المعلومات والأفكار من الوسائط المختلفة، كذلك أصدر الاتحاد الدولي لجمعيات المكتبات ومؤسستها عدة بيانات تؤكد على دور المكتبات في إتاحة المعلومات والمعرفة للجميع.

ولكن تبقى المشكلة في أرشفة مواقع الإنترنت في حرية إتاحة المواد التي تخضع لحماية قوانين الملكية الفكرية.

### ● حفظ المواقع للمستقبل

من المشكلات القانونية التي تواجه عملية أرشفة مواقع الإنترنت هي حفظ المواقع للاستخدام المستقبلي، فالمشكلة تتمثل في أن المكتبات مثل المتاحف تهدف إلى الحفاظ على التراث الفكري من الضياع.

### ١/٤ إجراءات أرشفة مواقع الإنترنت

تتطوى عملية أرشفة الإنترنت من الناحية التقنية على ثلاثة إجراءات رئيسية هي:

#### ■ الحصاد Harvesting

الحصاد عبارة عن عملية يستخدم فيها أحد البرامج ويقوم بجمع وتخزين محتوى مواقع الإنترنت وفقاً لما يحدده مستخدم البرنامج من محددات، وهي من العمليات المعقدة جداً لذلك فهي تتم بشكل آلي، ويوجد العديد من البرامج التي تستخدم في عملية الحصاد؛ وكان أولها البرنامج الخاص بمحرك البحث Alta Vista والسذي استخدم في اكتشاف مواقع الإنترنت، ظهر هذا البرنامج في منتصف التسعينيات تقريباً، بعد ذلك ظهرت البرامج التي تستخدم فقط في أرشفة مواقع الإنترنت وهي عبارة عن برامج طورت من خلال مشروعات أرشفة الإنترنت وبدأت في السويد أثناء تنفيذ مشروع أرشيف المواقع السويدية في ١٩٩٦، كذلك في أرشيف الإنترنت بالولايات المتحدة في نفس العام، ومع تزايد مشروعات أرشفة مواقع الإنترنت بدأ كل مشروع في تطوير أدواته الخاصة في حصاد المواقع، ففي المبادرة الأوروبية لأرشفة مواقع الإنترنت تم تطوير أداة حصاد الإنترنت في عام ٢٠٠٠ وهو متاح مجاناً.<sup>(٧)</sup>

وأياً كان البرنامج أو الأداة المستخدمة في حصاد مواقع الإنترنت، فهناك مجموعة من المواصفات الواجب توافرها في البرنامج، وكذلك بعض المتطلبات التي يجب أن يؤديها:

● السرعة في حفظ المواقع.

● أن يعمل البرنامج بشكل آلي في جميع المراحل.

● أن يتمكن من الوصول إلى عمق الموقع Deep Web<sup>(٨)</sup>

- في حالة عدم التمكن من حصاد عمق الموقع، على الأقل يكتشف البرنامج تلك الصفحات<sup>(٩)</sup>

#### ■ التخزين Storage

عملية التخزين من أهم متطلبات أرشفة الإنترنت، وبالطبع يتم التخزين على وسائط إلكترونية وعادةً ما تكون أقراص صلبة، وتتم عملية التخزين لحفظ ثلاثة مكونات رئيسية هي:

- قاعدة البيانات التي تتحكم في برنامج الحصاد، ويفضل أن تكون مبنية بنظام قوى لإدارة قواعد البيانات.
- مساحة للعمل الذي يسبق عملية الحفظ، مثل تجهيز الوثائق للحفظ، استخراج واصفات البيانات منها....
- المواقع التي ستحفظ في الأرشيف.<sup>(١٠)</sup>

#### ■ الإتاحة Access

عملية الإتاحة هي من الأهداف الأساسية لأي مشروع لأرشفة مواقع الإنترنت، فيجب أن يوفر الأرشيف وسيلة للبحث في محتوياته، وبالنظر إلى المشروعات الوطنية لأرشفة مواقع الإنترنت نلاحظ أن القليل منها يوفر إمكانية البحث، ويرجع ذلك إلى أن الكثير من المشروعات لا تزال قيد التنفيذ، ومن الإمكانيات التي يجب أن يوفرها الأرشيف في البحث هي إمكانية البحث التاريخي المحدد بفترة زمنية معينة حتى يتمكن المستخدم من استرجاع موقع ما كما كان في وقت سابق.

## ٢. مشروعات أرشفة مواقع الإنترنت في العالم

بدأت مشروعات أرشفة الإنترنت في عام ١٩٩٦ وقد كانت البداية في ثلاثة مشروعات تعتبر من أفضل الأرشيفات في الوقت الحالي، بدأت أرشفة الإنترنت في وقت واحد في المشروعات الثلاثة وهي: أرشفة الإنترنت العالمي **Internet Archive**، أرشفة الإنترنت للمواقع السويدية، أرشفة المواقع الأسترالية **PANDORA**. أرشفة الإنترنت الدولي هو أرشفة عام لمواقع الإنترنت في جميع دول العالم. أما الأرشيف السويدي والأسترالي فهما بداية لاتجاه أخذ في الانتشار لاحقاً في مجال مشروعات أرشفة مواقع الإنترنت وهو الاتجاه نحو إنشاء الأرشيفات الوطنية، ويمكن تقسيم الاتجاهات الدولية في إنشاء أرشفات الإنترنت إلى مايلي:

### ١ / ٢ اتجاه دولي

وهو يتمثل في أرشفة الإنترنت الدولي الذي أنشئ في سنة ١٩٩٦، وفي هذا الاتجاه يتم حصر جميع مواقع الإنترنت على مستوى العالم دون التقييد بأي معايير في الاختيار.

## ٢/٢ اتجاه إقليمي

بعد هذا الاتجاه خطوة أكثر تقدماً، حيث ظهرت مشروعات لإنشاء أرشيفات إقليمية لمواقع الانترنت الخاصة بمنطقة معينة، ومن أبرز الأمثلة على ذلك مشروع **NORDIC** وهو مشروع لإنشاء أرشيف لمواقع الانترنت في الدول الإسكندنافية التالية: الدنمارك، هولندا، أيسلندا، النرويج، فنلندا.

## ٢/٣ اتجاه وطني

وقد نما هذا الاتجاه نتيجة إدراك الدول لمدى الحاجة إلى حفظ تراثها الفكري المنشور على الانترنت، فجدد العديد من الدول شرعت في إنشاء أرشيف وطني لمواقعها على الانترنت، مثل: الدنمارك، جمهورية التشيك، بريطانيا، كندا، فنلندا، النرويج، اليونان، اليابان، سنغافورة، هولندا.

## ٢/٤ اتجاه موضوعي

ظهر هذا الاتجاه في سنة ٢٠٠١ وأخذ في التطور السريع ففي سنة ٢٠٠١ تقريباً نفذت منظمة أرشيف الانترنت بالتعاون مع "مكتبة الكونغرس" و "منظمة أرشيفوا الويب" **Web Archivists** مشروعاً لأرشفة مواقع الانتخابات الرئاسية في الولايات المتحدة الأمريكية، وتبعه مشروعاً آخر حول مواقع أحداث ١١ سبتمبر، انتقل هذا الاتجاه إلى أوروبا أيضاً حيث ظهر في هولندا مشروع **ARCHIPOL** وهو أرشيف لمواقع الأحزاب السياسية في هولندا، وغيرها من المشروعات التي وصل عددها حتى الآن ١٠ مشروعات تقريباً.

ويوضح الجدول رقم (١) المشروعات الحالية لأرشفة مواقع الانترنت في مختلف دول العالم.

جدول رقم (١) مشروعات أرشفة مواقع الانترنت

الدولة	اسم المشروع	تاريخ المشروع
الولايات المتحدة الأمريكية	1. أرشيف الإنترنت Internet Archive	

الدنمارك - السويد - النرويج - فنلندا - أيسلندا	1. أرشيف الدول الاسكندنافية	أرشيفات إقليمية
قارة أوروبا	2. مكتبات الإيداع الأوروبية المتشابهة	
أستراليا	1. أرشيف مواقع الإنترنت الأسترالية PANDORA	أرشيفات وطنية
بريطانيا	2. أرشيف المواقع الحكومية البريطانية	
السويد	3. أرشيف التراث الثقافي Kulturaw	
جمهورية التشيك	4. أرشيف الإنترنت	
الدنمارك	5. أرشيف الإنترنت	
اليونان	6. أرشيف مواقع الويب اليونانية	
اليابان	7. مشروع أرشيف الإنترنت WARP	
فنلندا	8. الأرشيف الإلكتروني التخيلي EVA	
كندا	9. مشروع المصادر الإلكترونية	
فرنسا	10. مشروع المكتبة الوطنية لأرشفة الإنترنت	
جزيرة تسمانيا	11. مشروع جزيرة تسمانيا الرقمية	
النمسا	12. الأرشيف للنمساوي على الخط المباشر	
النرويج	13. أرشيف الإنترنت	
هولندا	14. مشروع أرشفة الإنترنت	
نيوزيلاند	15. أرشفة مصادر الإنترنت	
الولايات المتحدة الأمريكية	1. أرشيف أحداث 11 سبتمبر	أرشيفات موضوعية
الولايات المتحدة الأمريكية	2. أرشيف انتخابات 2002 (الولايات المتحدة)	
الولايات المتحدة الأمريكية	3. أرشيف انتخابات الرئاسة الأمريكية 2000	
الولايات المتحدة الأمريكية	4. أرشيف انتخابات الرئاسة الأمريكية 2004	
سنغافورة - الولايات المتحدة الأمريكية	5. أرشيف كارثة تسونامي	

6. أرشيف الأحزاب السياسية ARCHIPOL	هولندا
7. أرشيف "الإنترنت والانتخابات"	سنغافورة - هولندا - الولايات المتحدة الأمريكية
8. نظام أرشفة مشروعات الطلبة	الولايات المتحدة الأمريكية
9. الأرشيف الرقمي للدراسات الصينية	الولايات المتحدة الأمريكية
10. الكونجرس الـ 107	مكتبة الكونجرس

### ٣. أرشيف الإنترنت الدولي Internet Archive

يواجه أى مستخدم للإنترنت عدة مشكلات وهي أنه عندما يدخل أحد مواقع الإنترنت تظهر له رسالة تفيد بأن هذه الصفحة غير موجودة، أو أن عنوان الموقع قد تغير، أو أن المعلومات التي كانت موجودة في الصفحة منذ فترة لم تعد موجودة، فكثيرا ما نجد معلومات مفيدة في أحد المواقع وعند الرجوع إلى الموقع نفاجأ بأن المعلومات التي اطلعنا عليها منذ فترة قد تغيرت وحل محلها معلومات أخرى، وبالتالي فهي مشكلة متعددة الأوجه، والسؤال هنا ما هو مصير تلك المعلومات الضائعة؟ الإجابة على تلك التساؤلات توجد عند فكرة أرشيف الإنترنت الذي يقوم بحفظ نسخة كاملة من مواقع الإنترنت على فترات مختلفة وإتاحتها للبحث بحيث يتمكن المستخدم من تصفح الموقع كاملا في أوقات سابقة وكما كان موجودا في الماضي، وسوف نتناول الفصل التعريف بأرشيف الإنترنت من حيث نشأته، وتاريخه، ومحتوياته، وطرق البحث والاسترجاع فيه، وحقوق الملكية الفكرية والجماليات التي يمكن استخدامها الأرشيف فيها، كما نتعرض لأرشيف الإنترنت المتوفر في مكتبة الإسكندرية.

ظهر أرشيف الإنترنت في سنة ١٩٩٦ وفكرة إنشاء أرشيف الإنترنت تعود إلى بروستر كاله **Brewster Kahle** أحد أشهر العلماء المتخصصين في مجال الإنترنت، حيث قام بإنشاء أرشيف الإنترنت سنة ١٩٩٦ بهدف حصر مواقع الإنترنت من جميع دول العالم والحفاظ عليها من الضياع، وذلك بحفظ نسخة من كل موقع على الإنترنت كل شهرين، واختارها في الأرشيف، وهذا العرض أنشأ كال منظمة غير ربحية تحمل نفس الاسم "أرشيف الإنترنت **Internet Archive**" وهو الرئيس الحالي لها، بالإضافة إلى أنه الرئيس الحالي أيضا لشركة أليكسا **Alexa** التي تتعاون مع أرشيف الإنترنت.

"أرشيف الإنترنت هو مستودع لحفظ مواقع الإنترنت يقوم بعملية أرشفة لمواقع الإنترنت من جميع دول العالم ويحفظها على فترات منتظمة أو غير منتظمة، وذلك باستخدام برامج تعمل بشكل آلي، بحيث يتمكن المستخدم



من تصفح أى موقع كما كان في أوقات سابقة، كما يحفظ مواد ثقافية أخرى مثل الأفلام والموسيقى وبرنامج الكمبيوتر"

أرشيف الإنترنت متاح مجاناً في الموقع التالي: [www.archive.org](http://www.archive.org)

### ٣/١ محتويات أرشيف الإنترنت

منذ أن بدأ أرشيف الإنترنت في العمل سنة ١٩٩٦ وهو يهدف إلى حفظ نسخ من مواقع الإنترنت، إلا أن الأمر تطور وبدأ في حفظ مواد أخرى بجانب مواقع الإنترنت مثل الأفلام السينمائية الأرشيفية، والموسيقى الحية، والوثائق والنصوص والآن أصبح أرشيف الإنترنت يتكون من تلك المجموعات الرئيسية، وهو كما يبدو واضحاً تقسماً وفقاً لنوع الوسيط الذي يتم حفظه واختزانه (مواقع إنترنت، أفلام سينمائية، موسيقى حية، وثائق ونصوص، برامج كمبيوتر، مصادر تعليمية)، وجدير بالذكر هنا أن كل هذه المجموعات مجانية وتتاح للعمامة بدون مقابل حيث إن منظمة أرشيف الإنترنت هيئة غير تجارية ولا تهدف للربح.

#### ■ مجموعة مواقع الإنترنت

وهي المجموعة الرئيسية في أرشيف الإنترنت وهي التي قامت فكرة إنشاء الأرشيف من أجلها، وهذه المجموعة تقوم على الفكرة الأساسية لأرشيف الإنترنت وهي حفظ مواقع الإنترنت في فترات مختلفة حتى يمكن إتاحتها للمستخدم كما كانت في أوقات سابقة، وتتكون مجموعة مواقع الإنترنت من مجموعات فرعية هي<sup>(١١)</sup>:

■ صفحات الإنترنت منذ ١٩٩٦ حتى الآن.

■ مواقع انتخابات الرئاسة الأمريكية لسنة ٢٠٠٠.

■ مواقع أحداث ١١ سبتمبر.<sup>(١٢)</sup>

■ مواقع انتخابات الولايات الأمريكية سنة ٢٠٠٢.

■ أرشيف كارثة إعصار كاترينا.

#### ■ مجموعة الأفلام

يقوم أرشيف الإنترنت بحفظ مجموعات كبيرة من الأفلام السينمائية والفيديو والمواد التلفزيونية، وكانت بداية تكوين هذه المجموعة عبارة عن حفظ مجموعة من الأفلام الأرشيفية حول الولايات المتحدة أنتجت منذ ١٩٢٧ وحتى الآن، إلا أن مجموعة الأفلام قد تطورت كثيراً الآن وأصبحت تضم ١٤ مجموعة فرعية، هذه المجموعة متاحة على موقع أرشيف الإنترنت مجاناً ويمكن الاطلاع عليها عن طريق التصفح حيث إنه لا تتوافر إمكانية البحث عنها، حيث ترتب المواد داخل كل مجموعة فرعية هجائياً وفقاً لعنوان المادة.<sup>(١٣)</sup>

وتتنوع مجموعات الأفلام بين برامج تليفزيونية أمريكية شهيرة، برامج إخبارية، أفلام سينمائية ووثائقية، وأفلام تتناول الحياة العامة الأمريكية، ومن الملاحظ أن الهدف من حصر هذه المجموعات في أرشيف الإنترنت هو هدف قومي أمريكي بالدرجة الأولى وليس فقط الحفاظ على هذه المواد من الضياع والاختفاء، وسوف نتناول فيما يلي التعريف بالمجموعات الفرعية التي تندرج تحت مجموعة الأفلام.

#### ■ مجموعة الموسيقى

وهي مجموعة للمواد السمعية، تضم مواد سمعية متنوعة بين مقطوعات موسيقية وأغاني وبرامج إذاعية، ومواد تعليمية، وهي بذلك تشبه مجموعة الأفلام ولكن الفرق في شكل المواد، تتاح هذه المجموعة على موقع أرشيف الإنترنت مجاناً ويمكن الإطلاع عليها عن طريق التصفح حيث إنه لا تتوفر إمكانية البحث عنها، وهي مرتبة تماماً مثل مجموعة الأفلام حيث ترتب المواد داخل كل مجموعة فرعية هجائياً وفقاً لعنوان المادة، وتشتمل هذه المجموعة على ٩ مجموعات فرعية<sup>(١٤)</sup>، وسوف نتناولها ببعض من التفصيل فيما يلي :

#### ■ مجموعة النصوص

وتتكون هذه المجموعة من وثائق ونصوص، وقد نشأت نتيجة للتعاون بين أرشيف الإنترنت وجهات أخرى لتنفيذ بعض المشروعات، هذه المواد نصوص لكتب متاحة في شكل رقمي، والهدف من تحويل هذه المواد في شكل رقمي - كما أعلن أرشيف الإنترنت - هو توفير إتاحة سهلة وبسيطة لمجموعة من النصوص الأرشيفية الغنية والمتنوعة، وهذه المجموعة مثل غيرها من مجموعات أرشيف الإنترنت متاحة مجاناً للاستخدام دون أى مقابل مادي، والمواد في كل مجموعة فرعية مرتبة هجائياً حسب العناوين<sup>(١٥)</sup>، وتتكون هذه المجموعة من ٦ مجموعات فرعية، منها:

#### ■ مجموعة وثائق اربانت ARPANET

■ مشروع المليون كتاب

■ مكتبة الأطفال

■ مشروع جوتنبرج

■ دليل تعليم الرقص

■ الكتب المفتوحة المصدر Open Source Books

■ برامج الكمبيوتر

هذه البرامج من الأرشيف أضيفت حديثا في سنة ٢٠٠٥، وهو أرشيف فرعى يهدف إلى حفظ وإتاحة الوصول إلى برامج الكمبيوتر النادرة والتي يصعب الحصول عليها الآن، وتشتمل هذه المجموعة على برامج مجانية متاحة للتنزيل مجانا وبشكل قانوني.

وتتميز هذه المجموعة بأنها توفر النسخ القديمة من برامج كمبيوتر معروفة الآن ويصدر منها حاليا نسخ حديثة، تبلغ عدد المواد الموجودة في هذه المجموعة ٣٤,٢٧٥ مادة.<sup>(١٦)</sup>

#### ■ المواد التعليمية مفتوحة المصدر

وهي المجموعات الحديثة المضافة حديثا إلى الأرشيف، وهي عبارة عن مجموعات من المواد التعليمية تضم: أدلة دراسية، مقررات دراسية، تدريبات، محاضرات مسجلة، وهذه المواد موجهة إلى الطلاب والمدرسين.<sup>(١٧)</sup>

#### ٣/٢ أدوات البحث والاسترجاع في أرشيف الإنترنت

بعد التعرف على الكم الهائل من المعلومات المتنوعة التي يخزنها أرشيف الإنترنت، السؤال الآن كيف يمكن استخدام كل هذه المعلومات؟ وكيف يمكن الوصول إليها؟

بالنسبة لمجموعات الأفلام الأرشيفية، والموسيقى الحية والنصوص فلا توجد وسيلة للاسترجاع حيث يمكن بسهولة كبيرة تصفح كل هذه المجموعات، أما بالنسبة لمجموعة مواقع الإنترنت فإن الأمر مختلف تماما حيث الكم الكبير من المواقع المخزنة والنسخ المتعددة من كل موقع وأيضا العدد الهائل من الصفحات في كل موقع والتي بلغ مجموعها أكثر من ١١ بليون صفحة، في ظل كل هذا من الطبيعي أن يوفر أرشيف الإنترنت الأدوات اللازمة لاسترجاع هذه المعلومات والوصول إليها، وقد طورت أداتين للبحث الأولى تسمى آلة الرجوع للخلف

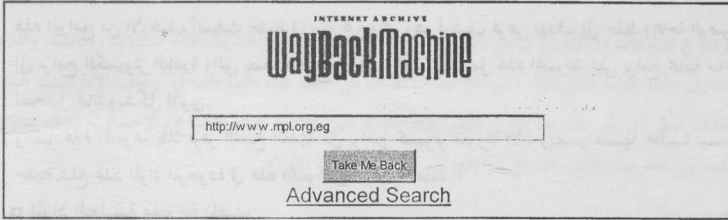
#### Wayback Machine ، والثانية أداة الاستدعاء Recall

#### ■ آلة الرجوع للخلف Wayback Machine

تم تطوير هذه الأداة بالتعاون مع أليكسا Alexa ويمكن من خلالها البحث عن طريق عنوان الموقع URL، بحيث يقوم المستخدم بكتابة عنوان الموقع حتى يمكنه الوصول إلى النسخ القديمة المحفوظة من الموقع، وعنوان الموقع هنا قد يكون العنوان الرئيسي للموقع، مثل: [www.cybrarians.info](http://www.cybrarians.info)، وقد يكون عنوان لصفحة فرعية داخل الموقع مثل: [www.cybrarians.info/about/team.htm](http://www.cybrarians.info/about/team.htm)، ولكن من الأفضل البحث بالعنوان الرئيسي للموقع إذا كان الهدف هو تصفح الموقع في الأرشيف والتعرف على محتوياته في وقت سابق، أما إذا كان الهدف هو الوصول إلى صفحة محددة داخل الموقع فيفضل استخدام الأسلوب الثاني وكتابة عنوان الصفحة الفرعية مباشرة<sup>(١٨)</sup> وتتيح آلة الرجوع للخلف نوعين هما: البحث بسيط والبحث

#### البحث البسيط

وكما يبدو واضحا من الشكل يتم إدخال العنوان المراد استرجاعه ثم البدء في البحث.



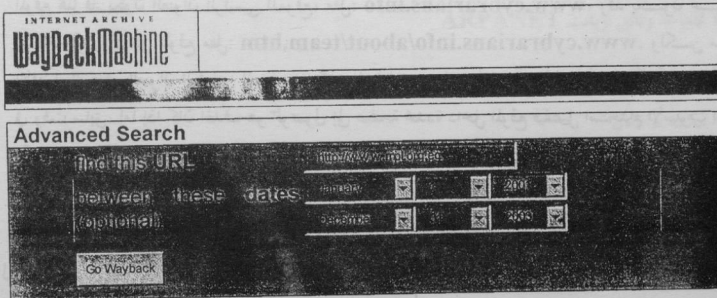
شكل رقم (١) البحث البسيط في أرشيف الإنترنت

## البحث المتقدم

وهو تقريبا يشبه البحث البسيط لكنه يضيف إمكانية تحديد فترة زمنية معينة نريد استرجاع الموقع فيها، بالإضافة إلى أنه يقدم بعض الخيارات التي يحددها المستخدم في نتيجة البحث<sup>(١٩)</sup>، منها على سبيل المثال:

- هل يرغب المستخدم في ظهور نتيجة البحث عن الموقع في Google , Yahoo ضمن نتيجة البحث في أرشيف الإنترنت؟ أم تظهر منفصلة؟
  - هل يرغب المستخدم في إظهار ضمن نتائج البحث صفحات إعادة التوجيه Redirection التي تحيل من عناوين إلى عناوين أخرى؟
  - ما هي أنواع الملفات التي يريد أن تظهر، نصوص ، فيديو ، صوت ؟ أم جميع الأنواع؟
  - إظهار النتائج المكررة أم لا؟ ويقصد هنا النسخ التي تحفظ لنفس الموقع الواحد لا أكثر من مرة.
  - إمكانية ضغط نتائج البحث، بحيث تظهر نسختان للموقع في نتيجة واحدة.
  - إمكانية تحويل نتائج البحث إلى شكل PDF، إلا أن هذه الإمكانية لا تزال تحت التجربة.
- و يوضح الشكل التالي نموذجاً للبحث المتقدم عن موقع مكتبة مبارك العامة، وفيه تم تحديد الفترة الزمنية من

١ يناير ٢٠٠١ حتى ٣١ ديسمبر ٢٠٠٣



شكل رقم (٢) البحث المتقدم في أرشيف الإنترنت

وأياً كان أسلوب البحث المتبع بسيطاً أو متقدماً، فإن نتيجة البحث تظهر مقسمة إلى أعمدة كل منها يعبر عن سنة محددة، و تحت كل سنة توجد النسخ الموجودة من الموقع في تلك السنة، وعلى المستخدم أن يضغط على النسخة التي يريد تصفحها، كما يوضح العدد الإجمالي للنتائج وعدد النتائج في كل سنة، و يوضح الشكل التالي نموذجاً لنتيجة البحث.

INTERNET ARCHIVE									
Wayback Machine									
Enter Web Address: <input type="text" value="http://www.mpl.org.eg"/> <input type="button" value="Take Me Back"/>									
Searched for <a href="http://www.mpl.org.eg">http://www.mpl.org.eg</a>								17 Results	
Search Results for Jan 01, 1996 - Jul 31, 2004									
1996	1997	1998	1999	2000	2001	2002	2003	2004	
0	0	0	0	0	5 pages	6 pages	6 pages	0 pages	
					Apr 02, 2001	Feb 16, 2002 *	Feb 19, 2003 *		
					Apr 05, 2001	May 24, 2002	Mar 25, 2003 *		
					Apr 19, 2001 *	Jun 05, 2002 *	Apr 22, 2003 *		
					May 15, 2001	Sep 28, 2002 *	May 26, 2003		
					Jul 20, 2001 *	Sep 29, 2002 *	Jun 18, 2003		
						Nov 21, 2002 *	Jun 20, 2003		

شكل رقم (٣) نتيجة البحث في أرشيف الإنترنت

#### □ أداة الاستدعاء Recall

و هذه الأداة تعتبر محرك بحث خاص بأرشيف الإنترنت، حيث يمكن من خلالها البحث في النص الكامل للمواقع المخزنة في الأرشيف، إلا أن هذه الأداة لا تزال تحت التطوير و لم تعمل بفاعلية بعد، و أسلوب البحث عن طريق هذه الأداة سهل للغاية و يشبه محركات البحث العادية، حيث يقوم المستخدم بكتابة الكلمة التي يريد البحث عنها و تحديد الفترة الزمنية التي يريد البحث فيها، و تتميز هذه الأداة في أنها لا تقدم نتائج البحث كما في الأرشيف بل تقدم بعض الإضافات التي تساعد المستخدم، حيث إنها تزود النتائج ببعض الإحصاءات حول هذا الموضوع، و حول تطوره الزمني و كل هذا يوضح بالرسوم البيانية، كما تقسم نتائج البحث إلى فئات موضوعية بمعنى أنها تظهر النتائج مقسمة وفقاً للموضوعات التي ظهر بها مصطلح البحث<sup>(٢٠)</sup>.

### Search the Archive

Searching 11,094,924,000 pages

شكل رقم (٤) نتيجة البحث في أرشيف الإنترنت

#### ٤- أرشيف مواقع الإنترنت الأسترالية PANDORA

يعد مشروع أرشيف المواقع الأسترالية من أوائل المشروعات في مجال أرشفة الإنترنت على مستوى العالم، بدأ المشروع في ١٩٩٦، والمشروع ليس مجرد أرشيف لمواقع الإنترنت الأسترالية ولكنه يعد دليلاً لمواقع الإنترنت الأسترالية حيث إنه يحرص المواقع ويصنفها موضوعياً، الأرشيف متاح للبحث على الإنترنت في العنوان التالي:

<http://pandora.nla.gov.au>

٤/١ نشأة الأرشيف

بدأ العمل في الأرشيف سنة ١٩٩٦ عن طريق المكتبة الوطنية الأسترالية وقد أطلقت عليه اسم **PANDORA (Preserving and Accessing Networked Documentary Resources of Australia)**، وهو يعني حفظ وإتاحة المصادر الوثائقية الأسترالية المتشابهة، ويبدو واضحاً أن المكتبة الوطنية الأسترالية تعتبر الإنترنت أحد مصادر التراث القومي للدولة حيث تنطلق عليها المطبوعات على الخط المباشر **On- Line Publication** كما أنها تضاف إلى البليوجرافية الوطنية، ويشارك مع المكتبة الوطنية في إعداد الأرشيف ٩ جهات أسترالية أخرى.

وهذا الأرشيف لا يعد فقط أرشيفاً لمواقع الإنترنت بل يعتبر أيضاً دليلاً بالمواقع الأسترالية حيث إنه يحرص المواقع الأسترالية ويقسمها موضوعياً إلى ١٥ قسماً رئيساً، وكل قسم ينقسم إلى أقسام فرعية كما نجد في أدلة و محركات البحث الشهيرة على الإنترنت.<sup>(٢٣)</sup>

٤/٢ معايير اختيار المواقع التي تحفظ في الأرشيف

وقد حدد الأرشيف معايير لاختيار المواقع، قسمت المعايير إلى ٣ فئات تلتخص في الآتي:<sup>(٢٣)</sup>،<sup>(٢٤)</sup>

##### ■ معايير عامة

- أن يكون أسترالي المحتوى
- تعدد الإصدارات
- المسؤولية والقيمة البحثية

##### ■ الفئات التي لها أولوية الاختيار

- المواد الحكومية الخاصة باستراليا
- مصادر المعاهد التعليمية المتوسطة
- أعمال المؤتمرات
- الدوريات الإلكترونية
- المواد التي تدخل ضمن خدمات التكشيف والاستخلاص

• المواد الموضوعية

• المواد الوطنية الأخرى

■ فئات أخرى

• وهي فئات أخرى من المواقع التي يحصرها الأرشيف منها: التقارير السنوية، الخرائط، المصادر التعليمية، الموسيقى، الأعمال الأدبية، المواقع الشخصية، الصحف، المواقع الدينية.

### ٤/٣ البحث و الاسترجاع في الأرشيف

وبالنسبة لاستخدام الأرشيف والبحث فيه، فإن الأرشيف يختلف تماما عن أرشيف الإنترنت فهو يوفر طريقتين للبحث في الأرشيف<sup>(٥)</sup>

• التصفح

يوفر الموقع إمكانية التصفح الموضوعي و الهجائي ل محتويات الأرشيف، بحيث يمكن للمستخدم الوصول إلى المواقع مباشرة وفقا لترتيبه الهجائي ، كما يمكنه الوصول إليه عبر تصنيفه الموضوعي كما يوضح الشكل

رقم (٥)

Browse Subjects		
Arts & Humanities	Health	News & Media
Business & Economy	History & Geography	Politics & Government
Computers & Internet	Indigenous Peoples	Science & Technology
Education	Juvenile	Society & Culture
Environment	Law & Criminology	Sports & Recreation

شكل رقم (٥) التصنيف الموضوعي في الأرشيف الأسترالي

• البحث البسيط والمتقدم

يوفر الأرشيف إمكانية البحث البسيط و المتقدم في محتوياته، بالنسبة للبحث البسيط فهو الشكل التقليدي للبحث باستخدام الكلمات المفتاحية، كما يوضحها الشكل التالي.

Search PANDORA	Browse Subjects
----------------	-----------------

شكل رقم (٦) شاشة البحث البسيط الأرشيف الأسترالي

و بالنسبة للبحث المتقدم، فإن الأرشيف يقدم إمكانات متميزة للبحث، حيث يسمح للمستخدم بتحديد: • التصنيف الموضوعي الذي يرغب في البحث عنه.

- الفترة الزمنية التي يرغبها.
  - البحث في عنوان الموقع فقط.
  - تحديد عدد النتائج التي تظهر في الصفحة الواحدة.
- و يوفر البحث المتقدم إمكانية البحث المحدد الموحد للمصدر URL و بالكلمات المفتاحية، كما يوضح الشكلين التاليين :

شكل رقم (٧) شاشة البحث المتقدم بالكلمات المفتاحية في الأرشيف الأسترالي

شكل رقم (٨) شاشة البحث المتقدم بالمحدد الموحد للمصدر URL في الأرشيف الأسترالي وبعد إجراء عملية البحث بأى من الطرق السابق ذكرها، فإن نتيجة البحث تظهر كالتالى:

library	Search	Advanced search	Search help
---------	--------	-----------------	-------------

Home > Advanced Search > Search Results

Results 1 - 10 of 989826 for library

[Australian library news](#)  
<http://pandora.nla.gov.au/tep/21843>

[Proston Public Library](#)  
<http://pandora.nla.gov.au/tep/33962>

[State Library of Queensland](#)  
<http://pandora.nla.gov.au/tep/39010>

[State Library of Victoria](#)  
<http://pandora.nla.gov.au/tep/44356>

شكل رقم (٩) نتيجة البحث في الأرشيف الأسترالي



وبالضغظ على أى نتيجة سيقدم الأرشيف عن كل موقع المعلومات التالية: عنوان الموقع، الجهة التي قامت بعملية الحفظ، تاريخ الحفظ.

#### Australian Music Radio Airplay Project

This title was selected for preservation by the National Film and Sound Archive: A division of the Australian Film Commission and has been archived once only. The publisher's site may provide more information:

Home page archived: 14 Nov 2002

شكل رقم (١٠) المعلومات المقدمة في نتيجة البحث في الأرشيف الأسترالي

#### ٤/٤ إحصائية حول الأرشيف

يشتمل الأرشيف على ٣,٣٠٠ موقع، ويبلغ العدد الإجمالي للملفات ١٨٧,٢١٠,٦٥١ ملف، الحجم الإجمالي للمواقع ٤٠٥ جيجا بايت، ومعدل الزيادة الشهرية يصل إلى ٢١ جيجا بايت في الشهر، وفيما يلي بيان مفصل بأعداد المواد المحفوظة لكل نوع من الملفات<sup>(٢٦)</sup>:

• نصوص HTML: ١٢٦,٥٨٧,٧٥٣

• نصوص TXT: ١,٥٢١,٦١٩

• صور: ٥٤,٠٤٣,٧٧٦

• برامج: ١,١١٧,٦٨٠

• ملفات PDF: ٣,٠٧١,٢٥٢

#### ٥. أرشيف المواقع الحكومية البريطانية

ظهر هذا المشروع نتيجة لتعاون الأرشيف الوطني البريطاني مع منظمة أرشيف الإنترنت بالولايات المتحدة الأمريكية، وهو عبارة عن أرشيف مجموعة مختارة من المواقع الحكومية البريطانية، أى أنه ليس أرشيفاً شاملاً لجميع المواقع البريطانية، ولكنه يحرص ٥٠ موقعاً حكومياً بريطانياً ويعمل على حفظها وإتاحتها بنفس الكيفية التي يعمل بها أرشيف الإنترنت، وقد بدأ في حفظ هذه المواقع منذ أغسطس ٢٠٠٣، وقد تعاقد الأرشيف الوطني البريطاني مع منظمة أرشيف الإنترنت على العمل في إنشاء هذا الأرشيف وإمداد بريطانيا بالخبرات والأدوات اللازمة لإنشاء هذا الأرشيف<sup>(٢٧)</sup>، الأرشيف متاح على الإنترنت في الموقع التالي:

<http://www.nationalarchives.gov.uk/preservation/webarchive>

وقد أنشأ الأرشيف الوطني البريطاني تجمع يحمل اسم " تجمع أرشيف مواقع الإنترنت البريطانية UK Web Archive Consortium، هذا التجمع مسئوليته تولى تنفيذ إنشاء الأرشيف الوطني للمواقع البريطانية، ويتكون هذا التجمع من ٦ جهات في بريطانيا يأتي على رأسها المكتبة الوطنية البريطانية، الأرشيف الوطني البريطاني، المكتبة الوطنية في ويلز، المكتبة الوطنية الأسكتلندية، وقد وضع التجمع جدولاً زمنياً لتنفيذ المشروع بحث يبدأ في يونيو ٢٠٠٤، وتبدأ عملية الأرشيف في يناير ٢٠٠٥، وفي يناير ٢٠٠٦ يتم عمل تقييم للعمل، وعلى أن ينتهي المشروع في يونيو ٢٠٠٦ (٢٨)

وفيما يتعلق بالجوانب التقنية في الأرشيف البريطاني، فإن عملية حفظ المواقع وفهرستها تتم عن طريق أحد برامج الزواحف **Crawlers** الخاصة الذي أعدته منظمة أرشيف الإنترنت، إلا أن هذا البرنامج مستخدم الآن بشكل مؤقت حيث قرر تجمع أرشيف مواقع الإنترنت البريطانية الاعتماد على البرنامج الذي طورته المكتبة الوطنية الأسترالية لإنشاء الأرشيف الوطني لمواقع الإنترنت الأسترالية، هذا البرنامج يسمى **PANDAS** (٢٩)

#### ١/٥ البحث والاسترجاع في الأرشيف

بالنسبة لإتاحة الأرشيف وكيفية البحث فيه، فهو مثل أرشيف الإنترنت يستخدم آلة الرجوع للخلف **Wayback Machine** في عملية البحث، إلا أنه بالطبع مخصص فقط للمواقع الحكومية البريطانية التي تم حفظها بشكل تجريبي وعددها ٥٠ موقعاً، ويقدم الأرشيف شاشة بسيطة للبحث فيقوم المستخدم بإدخال المحدد للمصدر **URL**، ولا يقدم الأرشيف إمكانية البحث الموضوعي أو في النص الكامل لمحتويات الأرشيف.

Search the UK Government Web Archive

You can search for a website by entering a web address and clicking on the Search button.

شكل رقم (١١) شاشة البحث البسيط الأرشيف البريطاني

ويوفر الأرشيف أيضاً إمكانية البحث المتقدم وهو تقريبا يشبه البحث البسيط لكنه يضيف إمكانية تحديد فترة زمنية معينة يريد المستخدم استرجاع الموقع فيها، بالإضافة إلى أنه يقدم بعض الخيارات التي يحددها المستخدم لتظهر في نتيجة البحث، على سبيل المثال:

- هل يرغب المستخدم في إظهار ضمن نتائج البحث صفحات إعادة التوجيه **Redirection** التي تحيل من صفحات إلى صفحات أخرى؟
- ما هي أنواع الملفات التي يريد أن تظهر، نصوص، فيديو، صوت؟ أم جميع الأنواع؟
- إظهار النتائج المكررة أم لا؟ ويقصد هنا النسخ التي تحفظ لنفس الموقع اليوم الواحد لأكثر من مرة.
- إمكانية ضغط نتائج البحث، بحيث تظهر نسختان للموقع في نتيجة واحدة.

شكل رقم (١٢) شاشة البحث المتقدم في الأرشيف البريطاني

وأيا كان أسلوب البحث المتبع بسيطاً أو متقدماً، فإن نتيجة البحث تظهر في شاشة مقسمة إلى أعمدة كل منها يعبر عن سنة محددة، وتحت كل سنة توجد النسخ الموجودة من الموقع في تلك السنة، وعلى المستخدم أن ينقر على النسخة التي يريد تصفحها، كما يوضح العدد الإجمالي وعدد النتائج في كل سنة، ويوضح الشكل التالي نموذجاً لنتيجة البحث.

شكل رقم (١٣) نتيجة البحث في الأرشيف البريطاني

## ٦. أرشيف التراث الثقافي- السويد Kulturarw'e

بدأت السويد في عام ١٩٩٦ العمل في مشروعها الوطني لإنشاء مواقع الإنترنت السويدية، وقد تولت القيام بهذا المشروع المكتبة الوطنية السويدية<sup>(٣٠)</sup>، أطلق على المشروع باللغة السويدية اسم Kulturarw'e وهو يعني أرشيف التراث الثقافي، وقد حدد المشروع لنفسه هدفا رئيسيا وهو حفظ وتجميع وإتاحة الوثائق الإلكترونية السويدية، والموقع الرسمي لهذا المشروع على الإنترنت هو: <http://www.kb.se/kw3/> الجهة المنفذة للمشروع هي المكتبة الوطنية للسويد، ولم تخصص المكتبة ميزانية إضافية للمشروع حيث يتم الصرف عليه من ميزانية المكتبة وقد تكلف المشروع منذ بدايته في عام ١٩٩٦ وحتى الآن حوالي ٢,٥ مليون يورو، علما بأن هذه الميزانية لا تشمل تكاليف أجهزة الحاسب أو التخزين.

### ٦/١ أهداف المشروع

- إنشاء نسخ متعددة من مواقع الإنترنت السويدية في أوقات متعددة.
- حفظ المواقع السويدية للمستقبل، والحفاظ عليها من الأندثار.
- توفير إتاحة سهلة لمجموعات الإنترنت السويدية.

### ٦/٢ معايير اختيار المواقع التي تحفظ في الأرشيف

لم يضع المشروع سياسة واضحة وصرحة لاختيار المواقع التي تحفظ في الأرشيف، إلا أن المشروع قد حدد المواد التي يسعى إلى حفظها في الأرشيف، ويمكن تحديد تلك المواد من حيث جانبيها؛ الأول: طبيعة الواقع، الثاني: طبيعة اسم النطاق. أولاً: وفقا لطبيعة المواقع.

وضع المشروع قاعدة عامة له وهي حفظ جميع مواقع الإنترنت السويدية ولكن مع إعطاء أولوية للفئات التالية من المواقع:

- مواقع الدوريات الإلكترونية السويدية
- مواقع المجموعات الإخبارية و القوائم البريدية السويدية
- مواقع الصحف و الجرائد السويدية
- مواقع المؤسسات و الجهات السويدية

### ثانياً: وفقا لطبيعة النطاق

حدد المشروع المواقع التي سيحصرها وهي التي لها اسم نطاق وطني سويدي وهو .se، ولكن واجه القائمون بالعمل مشكلات عديدة منها: وجود مواقع سويدية لا تستخدم اسم نطاق وطني يدل على انتماء الموقع للسويد، كذلك وجود جهات و مؤسسات غير سويدية وتستخدم اسم نطاق وطني سويدي وذلك لأنها

تقع داخل السويد ، وبشكل عام حدد المشروع القواعد التالية لاختيار المواقع وفقاً لامتداد أسماء النطاق الخاص بها:

- حصر جميع المواقع التي تنتهي بالامتداد الوطنى للسويد و هو .se.
- المستوى الأعلى لامتداد أسماء النطاق: .com, .org, .net, .edu.
- حصر المواقع التي تنتهي بالامتداد (nu) و هو امتداد شائع في السويد و يعنى أن المواقع بالسويدية (Now in Sweden)<sup>(٣١)</sup>

### ٦/٣ البحث و الاسترجاع في الأرشيف

حتى الآن الأرشيف غير متاح للاستخدام و البحث ، إلا أنه حدد بعض المواصفات والإمكانات التي يخطط المشروع تنفيذها فيما يتعلق بإمكانات بحث وتصفح الأرشيف ، وهي كالآتي :

- توفير إمكانات للبحث متوافقة مع ما هو متاح على الإنترنت من محركات بحث ، وذلك لضمان سهولة التعامل من قبل المستخدمين.

- توفير إمكانية البحث بعدد من العناصر كالموضوع والتاريخ.
- توفير إمكانية الربط بين أكثر من عنصر للبحث، أى تطبيق أسلوب البحث البوليني.
- توفير إمكانية تصفح الأرشيف.
- توفير إمكانية البحث في فترات زمنية محددة، وهذا يحقق الهدف الأساسي من أرشيف الإنترنت وهو الوصول إلى المواقع كما كانت في فترات سابقة<sup>(٣٢)</sup>.

### ٦/٤ إحصائيات حول مشروع الأرشيف

يتميز الأرشيف السويدي بنشر تقارير دورية حول المشروع، و أهم ما تشتمل عليه التقارير أرقام وإحصائيات حول الأرشيف ومحتوياته من المواقع، ومن هذه الإحصائيات التي تعبر عن الوضع الراهن للأرشيف، وفقاً لأحدث التقارير الصادرة عن المشروع بلغ إجمالي المواقع التي حصرها الأرشيف ٨٢٣,٠٠٠ موقعا حتى هذه المواقع الممتلئة في ٤٧٠ مليون ملف، وبلغ الحجم الإجمالي من مجموعات ١٨ تيرابايت، وقد توزعت أعداد المواقع وفقاً لامتداد اسم النطاق إلى ما يلي:

- .se : ٣٤٤,٠٠٠ موقعا
- .com : ٣٧٨,٠٠٠ موقعا
- .org : ٤٥,٠٠٠ موقعا
- .net : ٤٦,٠٠٠ موقعا
- .info : ٩٠٠ موقعا

• **.biz** : ٣٢٠ موقعا

• **Mil** : ١٤ موقعا

وفيما يتعلق بأنواع الملفات التي يحفظها الأرشيف، فهي كالتالي:

• **HTML** : ٢٦٣ مليون ملف.

• **صور** : ١٧٨ مليون.

• **برامج** : ١٣ مليون.

• **مواد صوتية** : ٢ مليون.

• **مواد مرئية** : ٣٠٦ ألف.

• **نصوص TXT** : ١١ مليون.

- 1 - **Buddharaju, Raju.** Recent developments in web archiving technologies. - Computer World - Vol. 12 Issue 3 (November 2005) - Accessed December 2, 2005 - Available at: <http://www.computerworld.com.sg/ShowPage.aspx?pagetype=2&articleid=2997&pubid=3&issueid=73>
- 2 - **Web Archiving Working Group.** Web Archiving Working Group Final Report. - University of Melbourne. - Updated June 2003 - Accessed December 5, 2005 - Available at: <http://www.unimelb.edu.au/web-archiving/docs/wawg.pdf>
- 3 - **Boudrez, Filip, and Eynde, Sofie Van Den.** DAVID: Archiving Websites - Leuven: Intersciplinair Centrum Voor Recht en Informatica, 2002. - P. 13. - Accessed November 21, 2005 - Available at: <http://www.antwerpen.be/david/eng/report05.pdf>
- 4 - **Day, Michael.** Collecting and Preserving the World Wide Web: a Feasibility Study Undertaken for the JISC and Wellcome Trust. - The Joint Information System Committee, 2003. - P. 8. - Accessed November 18, 2005. - Available at: [http://www.jisc.ac.uk/uploaded\\_documents/archiving\\_feasibility.pdf](http://www.jisc.ac.uk/uploaded_documents/archiving_feasibility.pdf)
- 5 - **WebHosting.Info :** The Internet Research Authority. - Accessed June 20, 2006 - Available at: [http://www.webhosting.info/registries/global\\_stats/](http://www.webhosting.info/registries/global_stats/)
- 6 - **Čolić, Alenka Kavčič.** Archiving the Web : some legal aspects - 68th IFLA Council and General Conference, Glasgow, August 18-24, 2002. - Accessed May 15, 2006 - Available at: <http://www.ifla.org/IV/ifla68/papers/116-163e.pdf>
- 7 - **Hakala, Juha.** Archiving the Web: European experiences. - Accessed December 2, 2005 - Available at: <http://www.lib.helsinki.fi/tietolinjat/0203/webarchive.html>
- 8 - **Deep Web :** مصطلح يطلق على صفحات الإنترنت المبنية باستخدام لغات البرمجة المتقدمة مثل : ASP, PHP ، وفيها يتم ظهور المعلومات في صفحة الويب عند فتحها من قبل المستخدم حيث تقوم باستدعاء المعلومات من قاعدة البيانات الخاصة بالموثق.
- 9 - **Masanès, Julien.** Towards Continuous Web Archiving: First Results and an Agenda for the Future. - D-Lib Magazine. - Volume 8 Number 12 (December 2002) - Accessed December 2, 2005 - Available at: <http://www.dlib.org/dlib/december02/masanès/12masanès.html>
- 10 - **Hakala, Juha.** Cited Opt.
- 11 - **Internet Archive.** - Accessed May 9, 2004. - Available at : <http://www.archive.org/web/web.php>
- 12 - **Schneider, Steven M. (et al)** Building Thematic Web Collections: Challenges and Experiences from the September 11 Web Archive and the Election 2002 Web Archive. - 3rd ECDL Workshop on Web Archives, Trondheim, Norway - 2003. - Accessed May 9, 2004. - Available at: <http://bibnum.bnf.fr/ecdl/2003/proceedings.php?f=schneider>
- 13 - **Internet Archive :** Moving Images Archive. - Accessed May 12 , 2004. - Available at : <http://www.archive.org/movies/movies.php>
- 14 - **Internet Archive :** Live Music Archive. - Accessed May 12 , 2004. - Available at : <http://www.archive.org/audio/etree.php>
- 15 - **Internet Archive Text Archive.** - Accessed May 12 , 2004. - Available at : <http://www.archive.org/texts/texts.php>
- 16 - **Internet Archive :** Software Archive. - Accessed October 22, 2005. - Available at : <http://www.archive.org/details/software>
- 17 - **Open Educational Resources.** - Accessed March 20, 2006. - Available at: <http://www.archive.org/details/education>
- 18 - **Koman, Richard.** How the Wayback Machine Works ? - 2003. - Accessed October 6, 2003. - Available at : <http://webservices.xml.com/pub/a/ws/2002/01/18/brewster.html>
- 19 - **Notess, Greg R.** The Wayback Machine: The Web's Archive. - Online. - Vol. 26, No. 2 (March/April 2002). - Accessed December 6, 2003. - Available at : <http://www.infotoday.com/online/mar02/OnTheNet.htm>
- 20 - **Help Page.** - 2004. - Accessed May 22, 2004. - Available at : <http://ia00406.archive.org/about.html>
- 21 - **Phillips, Margaret.** Archiving the Web: the National Collection of Australian Online Publications. - International Symposium on Web Archiving, Tokyo. - 2002. - Accessed June 20, 2004. - Available at: <http://www.nla.gov.au/nla/staffpaper/2002/philipps1.html>
- 22 - **Koerbin, Paul.** The role of the PANDORA Archive. - Australasian Sound Recording Association Conference. - 2004. - Accessed June 24, 2005. - Available at: <http://www.nla.gov.au/nla/staffpaper/2004/koerbin1.html>

- 25 - Australian Publications: Selection Guidelines for Archiving and Preservation by the National Library of Australia. - Last Updated August 25, 2003. - Accessed June 20, 2004. - Available at : <http://www.nla.gov.au/selectionguidelines.html>
- 26 - Margaret. What to collect and how to do it: the National Library of Australia's selective international Conference on Archiving Web Resources: Issues for Cultural Heritage National Library of Australia, Canberra, Australia. - 2004. - Accessed September 19, 2005. - Available at : <http://www.nla.gov.au/webarchiving/PhillipsMargaret.rtf>
- 25 - Cathro, Warwick and Webb, Colin and Whiting, Julie. Archiving the Web: The PANDORA Archive at the National Library of Australia. - Preserving the Present for the Future Web Archiving Conference, Copenhagen. - 2001. - Accessed June 24, 2004. - Available at : <http://www.nla.gov.au/nla/staffpaper/2001/cathro3.html>
- 26 - Koerbin, Paul. Report on the Crawl and Harvest of the Whole Australian Web Domain Undertaken during June and July 2005. - October 10, 2005. - Accessed December 12, 2005. - Available at : [http://pandora.nla.gov.au/documents/domain\\_harvest\\_report\\_public.pdf](http://pandora.nla.gov.au/documents/domain_harvest_report_public.pdf)
- 27 - The National Archive. - Accessed June 22, 2004. - Available at : <http://www.nationalarchives.gov.uk/preservation/webarchive/background.htm>
- 28 - Tuck, John. The British Library: Web Archiving Challenges and Initiatives. - International Conference on Archiving Web Resources: Issues for Cultural Heritage organisations, National Library of Australia, Canberra, Australia. - 2004. - Accessed September 19, 2005. - Available at : <http://www.nla.gov.au/webarchiving/TuckJohn.rtf>
- 29 - UK Web Archiving Consortium. - Accessed June 22, 2004. - Available at : <http://www.webarchive.org.uk/>
- 30 - Arvidson, Allan. Persson, Krister. Mannerheim, Johan. The Kulturaw3 Project : The Royal Swedish Web Archiw3e : An example of "complete" collection of web pages. - 2000. - Accessed July 06, 2004. - Available at : <http://www.ifa.org/IV/ifa66/papers/154-157e.htm>
- 31 - Arvidson, Allan. The collection of web pages at the Royal Library: the web heritage of Sweden. - 68th IFLA Council and General Conference, Glasgow, 2002. - Accessed July 08, 2004. - Available at : <http://www.ifa.org/IV/ifa68/papers/111-163e.pdf>
- 32 - Arvidson, Allan and Persson, Krister. Harvesting The Swedish Web Space. - ECDL Workshop, Darmstadt, Germany, 2001. - Accessed July 06, 2004. - Available at : [www.bnf.fr/pages/info/pro/ecdl/sweden/sld001.htm](http://www.bnf.fr/pages/info/pro/ecdl/sweden/sld001.htm)
- 33 - Kulturaw3 : Statistics. - Last Updated : February 14, 2005. - Accessed June 16, 2005. - Available at : <http://www.kb.se/kw3/ENG/Statistics.htm>