تقرير عن:

الميتاديتا لوصف واسترجاع مصادر المعلومات الإلكترونية على شبكة الإنترنت

إيناس فوزى
أخصائى مكتبات - بجامعة المنوفية

ملخص:

إن الميتاديتا هي الوسيلة الرئيسية لجعل البحث عن المصادر الإلكترونية على شبكة الإنترنت أكثر موصلية وكفاءة في تحديد المصدر أو المصادر وثيقة الصلة بما تم البحث عنه، فضلا عن دورها في تمكين أدوات البحث على الإنترنت من أن تكون أكثر إحكاما في ضبطها للمصادر المتاحة على الشبكة، فهي تساعد الباحث في إيجاد المصدر المناسب بسهولة من خلال ما تقدمه لأدوات البحث على الإنترنت عن بيانات متخصصة مهيكلة عن هذه المصادر، وبالإضافة إلى ذلك فإن الميتاديتا تعد مصدرا أساسيا لأخصائى المعلومات والمكتبات في بناء التسجيلات الببليوجرافية التي يعدها لوصف المصادر الإلكترونية على شبكة الإنترنت.

أسهم النمو الهائل للمصادر الإلكترونية المتاحة على شبكة الإنترنت في تعاظم مشكلة التعريف بهذه المصادر، ومن ثم الوصول إليها وإلى محتوياتها من المعلومات، وهو الهدف لكل من منتجي هذه المصادر والمستفيدين منها على حد سواء، لذلك كان من الطبيعي، لكي يتحقق هذا الهدف أن تتوافر مع كل مصدر من هذه المصادر بيانات تصفه وتحديد ملامحه وخصائصه بحيث يمكن لأدوات البحث، أو محركات البحث المختلفة على الشبكة اكتشاف هذه البيانات آليا وتجميعها داخل قواعد للبيانات تتاح للبحث عن المصادر.

ويطلق على هذه البيانات مصطلح «metadata»، وهناك العديد من المقابلات العربية لهذا المصطلح منها علي سبيل المثال: «البيانات عن البيانات»، «ما بعد البيانات»، «ما وراء البيانات».

وسوف نستخدم المقابل «البيانات عن البيانات» للتعبير عن مصطلح metadata لأنه الأكثر شيوعا.

ويمكن أن تعرف «البيانات عن البيانات» -met adata:

إجرائيا بأنها «البيانات الإلكترونية التي تصف خصائص المصادر الإلكترونية المتاحة على شبكة الإنترنت، وتحدد ملامح العلاقات بينها وتعزز من إمكانات اكتشافها، ومن ثم استخدامها بشكل كفء».

ويمكن توضيح مفهوم «البيانات عن البيانات»

بصورة أدق إذا أمكن تحديد طبيعة المصادر الإلكترونية المتاحة على شبكة الإنترنت وسماتها المميزة . فأي مصدر معلومات إلكتروني هو عبارة عن «وعاء رقمي : Digital item» يوضع ويعالج في صورة كيان مستقل بواسطة الإنسان أو الآلة، وهو قد يتشكل في وعاء واحد أو مجموعة متعددة من الأوعية، كما أن له - بغض النظر عن شكله أو نوعه - ثلاثة أوجه : -

الوجه الأول: المحتوى content

ويقصد به ما يحتوى عليه الوعاء من معلومات، وهو الوجه الداخلي للوعاء .

الوجه الثاني :السياق context

ويقصد به تحديد إجابات عن الأسئلة من ما، أين، لماذا، كيف؟ مما يتعلق بعملية إنشاء المصدر والتعامل معه، وهو الوجه الخارجي للوعاء .

الوجه الثالث: البنية structure

ويقصد به مجموعة الروابط المعيارية داخل الوعاء أو بين الوعاء وأوعية متعددة، ويمكن أن تتكون هذه الروابط أثناء إنشاء الوعاء أو بعد إنشائه .

وانطلاقا من هذا يمكن فهم البيانات عن البيانات : metadata على أنها ليست فقط البيانات التي تصف الوعاء أو تحقق ذاتيته، وتمثل محتواه الموضوعي، ولكنها أيضا توثق مسار الوعاء، وتحدد وظيفته وعلاقته بأوعية المعلومات الإلكترونية الأخرى، وكيفية إدارته والسيطرة عليه، وتقدم بيانات عن وضعية الحقوق الفكرية وشروط الإتاحة المتصلة به .

ويلاحظ أن «البيانات عن البيانات meta-data» : تقدم في شكل بيانات إلكترونية مهيكلة، أى في حقول وحقول فرعية ذات تيجان، وترفق بالمصدر الإلكتروني نفسه الذي تصفه - وهي تماثل في ذلك بيانات الفهرسة أثناء النشر بالنسبة للأوعية المطبوعة - حيث توضع في راس الوعاء Header، ولكن لا يراها المستفيد النهائي، ويقوم بإعدادها عادة منشئ المصدر نفسه أو تخلق إلكترونيا، ويكون ذلك عادة بلغات مثل : HTML, XML , SGML .

وهنا ينبغي الإشارة إلي أن بعض معايير الميتاديتا خاصة بترميز البيانات الوصفية فقط، في حين أن البعض الآخر يعمل علي ترميز كامل الوثيقة، بالإضافة إلي بيانات الوصف .

ويشير البعض إلى أن لغة الترميز القابلة للتوسع XML: Extensible Markup Language ستحل محل صيغة مارك . وتستخدم XML لوصف بنية ومحتوى الوثيقة الإلكترونية المتاحة علي الويب ، وتتميز هذه اللغة بالمرونة الشديدة - حيث تسمح لمستخدمها بإعداد الرموز التي يحتاجها، وتتفرع xmlفي حد ذاتها من اللغة المعيارية العامة لتهيئة النصوص SGML: Stan-dard Generalized Markup Language وهي تسمح بإنشاء نصوص وتوزيعها وتجهيزها في شكل متوج tagged text، ويتم من خلال SGML تحديد العناصر المختلفة للوثيقة بما في ذلك عناوين الفصول، والعناوين الفرعية،

والمستخلص، والحواشي، والفقرات، كما يتم تحديد العلاقات المنطقية التي تربط بين تلك العناصر.

ويمكن تمييز خمسة أنواع من «البيانات عن البيانات metadata» : وفقا لوظائفها كما يلي : -

النوع الأول: وصفي Descriptive

أي استخدام البيانات عن البيانات في تحديد ذاتية مصادر المعلومات الإلكترونية ووصفها.

النوع الثاني: إدارى Administrative

أى استخدام البيانات عن البيانات في إدارة مصادر المعلومات الإلكترونية والسيطرة عليها مثل البيانات المتعلقة بالتزويد، وحقوق الطبع والإنتاج، والتوثيق القانوني لمتطلبات الإتاحة، ونسخ الضبط والتمييز بين أوعية المعلومات الإلكترونية المتشابهة.

النوع الثالث: الحفظ Preservation

أي استخدام البيانات عن البيانات في إدارة عملية حفظ مصادر المعلومات الإلكترونية فهي توثق الحالة أو الشكل المادي للمصدر وتوثق أسلوب عمل حفظ النسخ المادية والرقمية لها.

النوع الرابع : فني Technical

أي استخدام البيانات عن البيانات في تحديد كيفية عمل وظائف النظام الذي يستخدمه المصدر في بيئته.

النوع الخامس: الاستخدام Use

أي استخدام البيانات عن البيانات في تحديد مستوي المصدر ونمط استخدامه.

وكان من نتيجة تعدد الممارسات المتعلقة باختيار ووضع وصياغة البيانات عن البيانات من جهة والنمو الهائل للمصادر الإلكترونية على شبكة الإنترنت التي تحتوى على البيانات عن البيانات من جهة ثانية أن دعت الكثير من الهيئات المعنية لعقد الاجتماعات وورش العمل والمؤتمرات لوضع تقنيات تحاول أن تضم عناصر البيانات عن البيانات في بناء واحد يمكن أن ترضى مختلف الاحتياجات ويمكن الالتزام به في وضع وصياغة هذه البيانات ، ولعل أهم هذه المحاولات الجهد الذي قام به مركز التحسيب المباشر للمكتبات OCLC، وشبكة معلومات المكتبات بانجلترا UKOLN اللذان أدركا أهمية وجود تقنين متفق عليه دوليا للبيانات عن البيانات، فانعقد الاجتماع الأول لمناقشة هذا الموضوع عام ١٩٩٥ بمشاركة العديد من المهتمين بالموضوع من خبراء الحاسبات والبرمجيات والنص الفائق وخبراء المكتبات الرقمية وأخصائي المكتبات والمعلومات، وكان عدد المشاركين ٥٢ مشاركا، وقد أسفر الاجتماع عن إجماع في الرأي علي وضع معيار دولي للبيانات عن البيانات أطلق عليه معيار DUBLIN CORE وقد سمي بذلك نسبة إلي مدينة دبلن بولاية أوهايو بالولايات المتحدة الأمريكية التي عقد بها الاجتماع.

ويمكن تقسيم مجموعة عناصر البيانات التي يضمها معيار Dublin Core إلى ثلاثة أقسام رئيسية تضم الـ ١٥ عنصرا وهي :

١ - المحتوى content

العنوان - الموضوع - الوصف - المصدر - اللغة ـ العلاقات - التغطية.

٢- المسئولية الفكرية intellectual property

المنشئ - الناشر - الموزع - حقوق الملكية .

٣- الفورية instantiation

التاريخ - النوع - الشكل - المعرف .

و تتميز مجموعة عناصر البيانات الخمسة عشر التي يضمها معيار DUBLIN CORE بما يلي : -

سهولة الإنشاء - سهولة التكشيف- القدرة علي تحقيق معدل استرجاع أفضل- قابلية المعالجة .

وقد توالت بعد ذلك الجهود والاجتماعات لوضع تقنيات للبيانات عن البيانات بحيث يمكننا الآن رصد الملاحظتين التاليتين :

الملاحظة الأولي:

وجود العديد من التقنيات التي تتميز بعضها ببساطته وعمومية تطبيقه على مختلف مصادر المعلومات الإلكترونية ,و البعض الآخر بتعقيده أو اقتصار تطبيقه علي أشكال أو نوعيات معينة من مصادر المعلومات الإلكترونية ، أو علي وظيفة من وظائف البيانات عن البيانات مثل : النصوص والمجموعات الأرشيفية والموسيقى . . .إلخ . مثال ذلك التقنيات الآتيه : -

1- Resource Description Framework (RDF)
2- Meta Content Framework (MCF)
3- Text Encoding Initiative (TEI)
4- Encoding Archival Description (EAD)
5- Governmental Information Locator Service (GILS)
6- Internet Anonymous FTP Archive (IAFA)
7- Resource Organization And Discovery in Subject based services (ROADS)
8- Summery Objective Interchange Format (SOIF)
9- Platform for Internet Content Selection (PICS)
10- Data elements for the exchange of cataloguing and metadata. ISO/DIS

الملاحظة الثانية:

كشفت واحدة من أحدث الدراسات[١] التي قدمت لنيل درجة الماجستير في المكتبات والمعلومات من جامعة القاهرة أن الوعي باستخدام تقنينات البيانات عن البيانات مازال قاصرا لدى منشئي المواقع المصرية على الإنترنت، وأن ما يؤكد ذلك ما يلي :

= أن ٨٨ , ٩٤٪ من المواقع المصرية وضع لها فقط تاج العنوان، فضلا عن أن من بينها ٣٢ , ٨٪ سجل عنوانها بطريقة خاطئة، ٠٤ , ٧٪ كان عنوانها يتكون من كلمة واحدة .

= أن ٦ , ٩٪ من هذه المواقع وضع لها تاج الكلمات المفتاحية .

= أن ٤ , ١٠٪ من هذه المواقع فقط وضع لها تاج الوصف .

(١) هشام فتحي أحمد مكي (٢٠٠١). مواقع البيانات على الإنترنت: دراسة نظرية وتطبيقية للمواقع المصرية توليفا وتحسيبا وإتاحه/ إعداد هشام فتحي أحمد مكي؛ إشراف سعد الهجرسى . أطروحة ماجستير . كلية الآداب. جامعة القاهرة.

وقد أعدت دراسة(*) لنيل درجة الماجستير في المكتبات والمعلومات من جامعة القاهرة (قيد البحث) عن معايير الميتاديتا لدراسة تقنينات الميتاديتا وأي من هذه التقنينات الأكثر ملاءمة في البيئة العربية والمقارنة بين الميتاديتا وتسجيلات مارك الحادي والعشرين.

المراجع

1- American library association (2000). Committee on cataloging: description and access taskforce on metadata final report June 2000. http://www.ala.org/alcts/organization/ccs/ccda/tfmeta6.html

2- Baker, Thomas (1996). Languages for Dublin core . D-lib magazine . dec 1998. http://www.dib.org/dlib/december98/12baker.html

3- Beckett, Dave (2000). Using Dublin core in xml http://purl.org/dc/docments/wd/dcmes-xml-2000714.html

4- Bekett, d.j (1996) Using Dublin core metadata : draft. http://www.hensa.ad.uk/pub/metadata/dc-encoding.html

5-Day,Michael (1997). Mapping Dublin core to unimarc : draft. http://www.ukoin.ac.uk/metadata/interoperability/dcunimarc.html

6-Doan , bich-lien (1998). Using metadata to improve organization and information retrieval on the www. http://orders.edrs.com/members/sp.cfm?an=ed427698

7- Hansen , preben (1997). User guidelines for Dublin core creation. http://www.sics.se/~preben/dc/dc_guige.html

8- Jessica , Milstead (1999). Metadata cataloging by any other name. Online , January/February , 1999 http:///www.onlineinc.com/onlineemag

9- Kasparova , n (2000) . Creation of the electronic recourses meta-database in Russia : problems and prospects . 66 th-IFLA council and general conference Jerusalem, Israel 13-18 august 2000. http://www.ifla.org/139-168e.html

(*) إيناس محمد فوزي أبو النور (٢٠٠١). تقنينات البيانات الخلفية لوصف واسترجاع المصادر الإلكترونية علي شبكة الإنترنت/ إعداد إيناس محمد فوزي أبو النور ؛ إشراف مصطفي حسام الدين. أطروحة ماجستير. كلية الآداب. جامعة القاهرة. (قيد البحث).