المحاضرة 02: من البحث الوثائقي الى استرجاع المعلومات

يشير" لانكستر "في كتابه **أساسيات استرجاع المعلومات** إلى أن استرجاع المعلومات[1] هو النشاط الأساسي لما تقوم به مرافق المعلومات، من المكتبات ومراكز المعلومات والتوثيق، ومنتجي بنوك وقواعد البيانات، سواء المنشورة في شكل مطبوع أو الإلكترونية ، وأية نوعية أخرى من المرافق التي تقدم مصادر المعلومات لمجتمع المستفيدين

1-تعريف نظم استرجاع المعلومات:

تعددت تعاريف نظم استرجاع المعلومات من طرف الباحثين في علوم المكتبات وعلوم الحاسوب ابتداء من خمسينات القرن العشرون وذلك لاختلاف وجهات النظر واستخدام هذا المصطلح في عدة تخصصات كما ان تطور تكنولوجيا المعلومات وظهور الانترنت عملت على توسع وزيادة هذه التعاريف نذكر منها :

-مجموعة من الاجراءات المميكنة عادة تستخدم في الرجوع الى البيانات التي تحتويها الوثائق وتكشيف تللك البيانات واختزانها بطريقة يمكن استعادتها عند الطلب.[2]

-كما عرف نظام استرجاع المعلومات بأنه مجموعة من النظم الفرعية والمتمثلة في الاتي:[3]

- النظام الفرعي الخاص باختيار الوثائق.
- النظام الفرعي الخاص بالتكشيف.
- النظام الفرعي الخاص باللغة.
- النظام الفرعي الخاص بالبحث.
- النظام الفرعي الخاص بالتفاعل بين المستفيد والنظام.
- النظام الفرعي الخاص بالمضاهاة.

وتتفاعل هذه النظم مع بعضها البعض لتحقيق الهدف من انشاء النظام وهو استرجاع المعلومات.

فنظام استرجاع المعلومات هو نظام يتعامل مع معلومات وسيطة بين المستفيد والمعلومات النهائية التي يسعى اليها فهو نظام لا يزيد الحالة المعرفية للمستفيد منه ،ولكن يحيط المستفيد علما بوجود معلومات عن الوثائق التي تتعلق بموضوع استفساره.

---

[1] -هناك مصطلح مقارب لاسترجاع المعلومات هو البحث الوثائقي الذي يعرف بانه:
« Action, méthodes et procédure ayant pour objet de trouver dans de fonds documentaires les références des documents pertinent ». Chevalier, Max. Usages et recherche d'information : Habilitation a diriger des recherches : informatique : université de Toulouse 3 :2011.p25

[2] -Muddamalle, M R . **Natural language versus controlled vocabulary in information retrieval: a case study in soil mechanics**. Journal of the American Society for Information Science , vol. 49 , n.10(Aug 1998) . P 881-887.

[3] -حسب الله، سيد.الشامي، احمد محمد. **الموسوعة العربية لمصطلحات علوم المكتبات والمعلومات والحاسبات: انجليزي –عربي.** القاهرة:المكتبة الأكاديمية،2001. ص138.

فاسترجاع المعلومات هو العملية او الطريقة التي تمكن المستخدم من تحويل حاجته للمعلومات الى قائمة فعلية من الوثائق الموجودة في المخزن والتي تحتوي المعلومات المفيدة للمستخدم وتلبي رغباته.

2-أنواع نظم استرجاع المعلومات: تقسم نظم استرجاع المعلومات الى الاقسام التالية:[4]

1-نظم استرجاع المعلومات: تعني استرجاع التسجيلات الببليوغرافية للوثائق.

2-نظم استرجاع الحقائق: تتمثل في النظم التي تعني بالرد على الاستفسارات فتقدم للمستفيد بيانات احصائية او جوانب فيزيائية (مثل بيانات الخواص الحرارية الفيزيائية للمواد).

3-نظم استرجاع الوثائق: وهي النظم التي تختزن النصوص الكاملة لمجموعة وثائق يتم استرجاعها من خلال البيانات الببليوغرافية والكلمات المفتاحية.

3-الصعوبات التي تواجهها نظم استرجاع المعلومات:

يرى بعض الباحثين ان مشكلة استرجاع المعلومات اكثر تعقدا مما تبدو لأول وهلة وتوصف عادة بانها غامضة غير محددة المعالم وترجيحية واحتمالية ويبدو ان الاتجاه السائد الان هو التركيز على جانب المخرجات في الانشطة (التعبير عن الحاجة الى المعلومات والاستفسارات)اكثر من التركيز على جانب المدخلات (التحليل الوثائقي لمصادر المعلومات) اعتقادا ان جانب المخرجات اكثر تعقدا.

وتواجه نظم استرجاع المعلومات صعوبات عدة تتمثل في:

ا-المدى:في قواعد المعلومات الكبيرة مثل فهارس المكتبات المحسبة وقواعد البيانات الببليوغرافية على الخط المباشر قد يتم تزويد المستفيد بكم كبير من النتائج في حين يكون في حاجة الى بيانات وثيقة واحدة او بعض الوثائق القليلة الكافية للإجابة على استفساره.[5]

ب-السمات المتعددة: يعتمد الكثير من نظم استرجاع المعلومات على عدد من السمات المختلفة للمواد حيث يمكن للمستفيد استرجاع الوثائق باستخدام الكلمات المفتاحية او المؤلف او تاريخ النشر وغير ذلك بغرض تلبية الاحتياجات المختلفة للمستفيدين وتحقيق القدرة على استرجاع الوثيقة الواحدة من خلال اكثر من سمة واحدة كالتاريخ والموضوع والمؤلف والعنوان وفقا للحاجة وقد يصل الامر الى جعل نظام الاسترجاع غير اقتصادي وغير عملي.

ج-التحديد:

ليست كل الأشياء على الدرجة نفسها من السهولة لتحديدها وضبط الكلمات المدخلة في نظام المعلومات لاسترجاع وثائق عنها، فهناك موضوعات محددة يمكن وصفها بسهولة وهناك موضوعات اكثر تعقيدا وصعوبة لابد من مراعاة جانب التحديد من قبل كل من الباحث والمكشف على حد سواء بل يجب على

[4] -Rowley, Jennifer . **Organizing Knowledge: an Introduction to Information retrieval**. 2nd ed . USA :Ashgate, 1992.p29

[5] - لانكستر،ف. وورنر،ا.ج. **اساسيات استرجاع المعلومات** ،ترجمة حشمت قاسم . الرياض: مكتبة الملك فهد الوطنية،1997. ص -

المكشف ان يلتزم بمستوى تحديد السمات واستخدام نفس الواصفات لوصف المحتوى للوثائق في نظام استرجاع المعلومات.[6]

4-لغات نظم استرجاع المعلومات:

تستخدم لغة النظام في مرحلة التكشيف لتحديد المصطلحات الملائمة لموضوع او موضوعات الوثيقة التي يتم تكشيفها كما تستخدم ايضا في مرحلة الاسترجاع عند صياغة استراتيجيات البحث الملائمة حيث يتم مضاهاة المصطلحات المختارة كمصطلحات بحث مع مصطلحات التكشيف المستخدمة في النظام.

ويمكن تقسيم لغات نظم استرجاع المعلومات الى قسمين :

4-1-اللغات الطبيعية (غير المقيدة):

يستخدم المكشف نفس اللغة التي استخدمها المؤلف في التعبير عن الافكار والمفاهيم في نص الوثيقة او عنوانها دون الحاجة الى الاستعانة بأدوات خارجية لاختيار مصطلحات صناعية مقننة.

ويطلق على التكشيف المعتمد على اللغات الطبيعية التكشيف الاشتقاقي لان المصطلحات تكون مشتقة او مقتبسة من النص نفسه.

ويرجع الى الباحث Mortimer Taube الريادة في هذا المجال حيث عبر عن إمكانية التعبير عن المحتوى الموضوعي للوثيقة بمصطلحات أحادية يتم اشتقاقها من الوثيقة نفسها دون الحاجة إلى تقييد تلك المصطلحات في ملف استنادي موضوعي.[7]

4-2-اللغة الاصطناعية(المقيدة):

وهي مصطلحات ورموز يستخرجها المكشف من ادوات خارجية وقد تكون تلك الادوات عبارة عن قوائم رؤوس الموضوعات او المكانز او خطط تصنيف ويتقيد المكشف باستخدامها عند التعبير عن المحتوى الموضوعي للوثائق ويسمى التكشيف المعتمد على اللغة المقيدة بالتكشيف المقيد او التعييني.[8]

4-3-أدوات اللغة المقيدة:

هي لغات صناعية مكونة من مفاهيم و مصطلحات تربط بينها عدة علاقات تهدف الى تمثيل المفاهيم الموجودة في الوثائق و جمل المساءلة .و هي تكون رابط بين لغة الوثائق و لغة جملة المساءلة التي يعبر بها المستفيد عن حاجياته ،تقسم اللغات التوثيقية الى صنفين كبيرين لغات ذات بنية هرمية مثل التصانيف و لغات ذات بنية تحليلية مثل المكانز.

---

[6] - لانكستر،ف. وورنر،ا.ج.**المرجع نفسه** . ص24

[7] - Lancaster , F.W. **Vocabulary Control for Information Retrieval**. 2nd ed . Arlington, Virginia : Information Resources Press , 1986 . p161.

[8] - محمد فتحي عبد الهادي، يسرسة محمد عبد الحليم زايد. **المرجع نفسه** .ص 112

ا- قوائم رؤوس الموضوعات:

هي قوائم مقننة تحصر المصطلحات أو رؤوس الموضوعات و تساعد في اختيار المناسب منها لوصف المحتوى الموضوعي للوثائق، و تشمل قوائم رؤوس الموضوعات على المصطلحات التي يمكن استخدامها و على الاحالات التي يجب عملها في الفهرس ، و تعتمد المكتبة في اختيار رؤوس الموضوعات و صياغتها على قواعد محددة و واضحة[9] .

و قوائم رؤوس الموضوعات القياسية المستخدمة في المكتبات تقع في مجموعتين:

أ-قوائم رؤوس موضوعات عامة للمكتبات العامة.

ب-قوائم رؤوس موضوعات متخصصة.

و قد بدأت قوائم رؤوس الموضوعات في الظهور ابتداء من سنة 1876 مع قواعد كتر للفهرسة الموضوعية و الوصفية ، و أول رؤوس موضوعات مقيسة هي قائمة جمعية المكتبات الأمريكية لرؤوس الموضوعات (A.L.A List of Subject Heading)الذي نشر أول مرة سنة 1895[10] .

ب- المكانز: يعرف المكنز على انه قوائم المصطلحات أو الواصفات التي تستخدم لتقنين وصف المفاهيم التي تشمل عليها الوثائق، و تبين المكانز العلاقات الهرمية بين المصطلحات و هي بذلك لغة صناعية تستخدم في نظم استرجاع و اختزان المعلومات[11].

كلمة مكنز thesaurus كلمة مشتقة من اليونانية و هي تعنى مستودعا للمعرفة ، ظهرت الكلمة في اللغة الانكليزية سنة 1736 و كانت تعني قاموس أو موسوعة.

وتطور استخدام كلمة (thesaurus) حوالي نصف القرن 19، فظهرت في قاموس اللغة الانجليزية للكاتبPeter Mark Roget سماه « Thesaurus of English Word and phrases » حيث بنى قاموسه الجديد على ترتيب الكلمات ليس الفبائيا لكن حسب الأفكار[12].

أما أول مكنز ظهر فيعود لسنة 1959 في الولايات المتحدة الأمريكية « thesaurus of ASTIA descripteurs» في وزارة الدفاع ثم طور و عرف باسم مكنز TEST (thesaurus of engineering and scientifique termes) و يحوي حوالي 25.000 واصفة .

والمكنز عبارة عن مجموعة مصطلحات و كلمات تستخدم في تصنيف و تكشيف و تخزين و استرجاع المعلومات و تكون عادة كلمات مفتاحية أو مصطلحات واصفة أو عناصر بيانات و قد تكون وفق ترتيب هجائي أو تركيب هرمي و لا يقتصر دور المكنز في اسهامه في عملية التكشيف و التحليل الموضوعي

---

[9]-محمد فتحي عبد الهادي. **الفهرسة الموضوعية: دراسة في رؤوس الموضوعات**. جدة: دار الشروق،1981.ص192

[10] -محمد فتحي عبد الهادي.**المرجع السابق** .ص 192

[11] -Accart ,Jp ,Rethy,Mp .**Le métier de documentaliste** . paris : Edition du cercle de la librairie.1999,p294

[12]- محمد فتحي عبد الهادي. **المرجع نفسه** .ص 107

للوثائق و لكنه يتجاوز ذلك الى اسهامه في عملية البحث بمساعدة الباحث على اختيار مصطلحات البحث المقننة و بالتالي بناء استراتيجية البحث بشكل أفضل[13].

لقد تم استخدام الحاسوب الآلي في تسيير و انتاج و صيانة المكانز، كما تم اضافته الى برمجيات البحث الوثائقي فأصبحت تسمى برمجيات التسيير و البحث الوثائقي LGRD كما ظهرت برمجيات انشاء و صيانة المكانز عرفت باسم (standalone) .[14]

و يمكن تلخيص الوظائف الأساسية للمكنز في:

1-انها تتيح المجال للمكشف في تمثيل المادة الموضوعية المحتواة في الوثائق بطرقة ثابتة موحدة ، حيث تسمح للمكشف أن يصف المعلومات التي تحتويها الوثائق بطريقة أكثر اكتمالا و على مستويات مختلفة من العمومية و من وجهات نظر فنية متعددة.

2-أنها تمد الباحث بالوسائل التي يمكن من خلالها أن يعدل استراتيجية البحث من أجل تحقيق نسبة استرجاع عالية ، فضلا عن الدقة و ذلك وفقا لظروف و متطلبات البحث و الدراسة .

---

[13] - Erlos, Frédéric. **Thésaurus et accès à l'information référentiels terminologiques adaptables au contexte : l'exemple d'un système de recherche d'informations dans une grande entreprise**. 7es Journées internationales d'Analyse statistique des Données Textuelles (JADT).(En line)www.cavi.univ-paris3.fr/lexicometrica/jadt/jadt2004/tocJADT2004.htm . visite le (8/9/2011 à 16.00)

[14] - Dalbin ,Sylvie. **Thesaurus et informatique documentaire : partenaires de toujours ?** Documentaliste : Sciences de l'information 2007, vol. 44, n° 1.p42-55