الإنترنت والأطر البحثية
فـى استرجـــاع المعلـــومات

د. ناريمان إسماعيل متولى
أستاذ علم المكتبات والمعلومات المساعد
كلية الآداب – جامعة الإسكندرية

اولاً: مقدمة :

تشير مراجعات تقدم أطر البحث فى مجال استرجاع المعلومات (Ellis, D, 1996) إلى وجود مدخلين لهذه الأطر التى سيطرت على مجال استرجاع المعلومات ، أما المدخل المفهومى الأول والذى كان سائداً منذ الخمسينيات والستينيات فقد أطلق عليه المدخل النموذجى أو المثالى أو الأصلى Archetypal وهو المدخل الذى يعتمد على الاختبارات المقارنة لنظم ولغات التكشيف ، أما المدخل المفهومى لثانى فيعتمد على الاهتمامات المعرفية Cognitive التى كانت سائدة من منتصف السبعينيات .

وعلى الرغم من صحة هذا التحليل فى نهاية التسعينيات من القرن العشرين ، إلا أن هناك تغييرات واضحة فى عالم بحوث استرجاع المعلومات ، والتى تمت منذ ظهور الإنترنت وإحدى خدماتها الأكثر أهمية وهى الشبكة العالمية العنكبوتية (Web) فضلاً عن دخول الانترانت Intranet والتى يمكن تعريفها بأنها انترنت داخلية Internal Internet تستعين بفاحص الوب Web Browser ومتكاملة مع البنية المعلوماتية التحتية للمؤسسة ، وإذا كانت الانترنت تزودنا بإتاحة محتوياتها لأي عضو من الجمهور المستخدم ، فإن الإنترانت تضبط اتاحتها لمحتوياتها بحيث تسمح فقط للمخولين من الأعضاء بالإفادة منها ، وتحتوى الانترانت على صفحات الوب وقواعد بيانات وثائقية من المعلومات التى تتوفر فى القائم بالخدمات Server خلف جدران الانترنت .

ونحن نخوض تجربة مثيرة فى عالم المعلومات بعد ظهور الانترنت عن طريق فيضان المعلومات الإلكترونية المتاحة فى الإنترنت ، بما تشمله من معلومات محلية وأقراص مكتنزة CD-Rom وقواعد بيانات على الخط المباشر ، أى أن الإنترنت تزودنا بالمعلومات المحليةو التقليدية (مثل الديالوج) ومواقع الوب التى تحتوى على مستخلصات أوباك OPAC والوثائق ذات النصوص الكاملة Full-Text وأوعية الوسائط المتعددة Multimedia لمصادر المعلومات ..

وكل هذه التغييرات السريعة تبرز قضايا جديدة ذات علاقات بحثية بالمجالات المختلفة لاسترجاع المعلومات.

وعلى سبيل المثال فقد قامت مجلة بحوث علم المكتبات والمعلومات بإعداد جدول البحوث المقترحة خلال بدايات القرن الحادى والعشرين (Editorial, 1997) بناء على إجابات أعضاء مجلس تحرير المجلة وتركزت هذه فى عشرة مجالات لها علاقة بطريقة أو أخرى ببحوث استرجاع المعلومات والإنترنت وهى:

١ – جوانب استرجاع المعلومات ذات العلاقة بالإنترنت والوب (بما فى ذلك جودة الخدمة).

٢ – محركات البحث والألجورزم (الخوارزميات) Search Engines and Algorithms .

٣ – دور الفهرسة والتصنيف فى الضبط الببليوجرافى لمصادر الإنترنت .

٤ – تفاعل المستفيد مع الحاسب Use-Interface شاملاً التفاعل العام والاستعانة باللغة الطبيعية .

٥ – تحليل متطلبات واحتياجات المستفيد وكيفية جذبهم للمكتبة .

٦ – سلوك البحث عن المعلومات والاتجاه المعرفى Cognitive Approach .

٧ – المكتبات والمعلومات الرقمية والبحث فى إنشاء واسترجاع واستخدام الأشياء الرقمية غير النصوص .

٨ – تلخيص وتمثيل المعلومات وفحص أنواع القياسات ومدى ملاءمتها للاسترجاع .

٩ – المعلومات المرئية Visual Information .

١٠ – استرجاع المعلومات الذكى بما فى ذلك العوامل الذكية Intelligent Agents .

وقد لوحظت هذه المجالات فى تقارير (Text Retrieval Conference 1-5) TREC ومشروعات أوكابى (Okapi and Information Retrieval Research 1997) OKAPI (Smeaton, A. F., 1997) .

وفى محاولة للتعرف الأكثر قرباً بالمجالات ذات الارتباط بين الإنترنت واسترجاع المعلومات اظهر البحث الخاص بقاعدة بيانات ليزا LISA وفحص الأعداد الحديثة من دوريات المكتبات والمعلومات (Chowdhury, G., 1999: 210) المجالات التالية ذات التقارب الشديد مع القائمة السابقة (وإن اختلفت فى التعبير عن المحتوى مثل الضبط الببليوجرافى على الإنترنت وتنظيم المعلومات على الوب)

١ – محركات البحث .

٢ – تقييم الاسترجاع Evaluation .

٣ – الثقة فى معلومات الوب .

٤ – التقابل مع المستفيد Interface .

٥ – دراسات المستفيدين .

٦ – تنظيم المعلومات على الوب .

٧ – ضبط المصطلحات .

٨ – مخرجات البحث من الوب .

٩ – عوامل البحث الذكية .

١٠ – الوب ومقارنته بقواعد البيانات التقليدية (on-line and CD-Rom)

ثانيا : مشكلة الدراسة وتساؤلاتها والدراسات السابقة ومنهج البحث :

إذا كانت الباحثة قد تناولت فى دراستها السابقة (ناريمان متولى ، يناير ٢٠٠٠ م) معظم البحوث الواردة فى القائمة الثانية أعلاه ذات العلاقة بالإنترنت فستركز مشكلة الدراسة الحالية فى التعرف على التطورات الأحدث بالنسبة لبعض الجوانب السابق دراستها مع إضافة بعض الجوانب الأخرى الواردة فى القائمة الأولى ويمكن أن تتبلور مشكلة الدراسة فى التساؤلات التالية :

١ - ما الذى نعنيه بالإطار Paradigm وما هى مقومات كل من المداخل البحثية النموذجية Archetypal والمعرفية Cognitive والتى تحقق صفات الهوية الإطارية لمجال استرجاع المعلومات .

٢ - ما هى مشكلات القياس ودوره فى أطر بحوث استرجاع المعلومات .

٣ - ما هى مستحدثات محركات البحث ومدى استخدامها بذاتها فى الاسترجاع أو مع عوامل البحث الذكية لحل المشكلات التى تعانى من البحث البولينى .

٤ - ما هو موقع الانترنت المقارن بالنسبة لخدمات الاسترجاع الرقمية بالمكتبات الأكاديمية .

أما بالنسبة للدراسات السابقة ومنهج البحث :

لا تتوفر دراسات باللغة العربية عن الأطر البحثية لاسترجاع المعلومات وإن كانت هناك بعض الدراسات ذات الصفة العامة عن الأطر التاريخية والاجتماعية والطبيعية والمعرفية لعلم المعلومات ، كعلم متعدد الارتباطات الموضوعية والنظرية (أحمد بدر ٢٠٠٢) ولكنها لا تتناول أطر بحوث استرجاع المعلومات ، كما يحفل الإنتاج الفكرى العربى الحديث المتمثل فى القوائم الببليوجرافية الشاملة للإنتاج الفكرى العربى ببعض المعالجات بين الإنترنت واسترجاع المعلومات ، ولكنها لا تتضمن أطر بحوث استرجاع المعلومات ، وأخيراً فمنهج الدراسة المتبع فى الدراسة الحالية هو منهج الوصف والتحليل المتمثل فى استقراء الحقائق من الإنتاج الفكرى وتجميع الأدلة اللازمة للإجابة على التساؤلات التى وضعتها الباحثة .

ثالثا : مفهوم الإطار ودوره في بحوث استرجاع المعلومات :

يضم الإنتاج الفكرى لبحوث استرجاع المعلومات تسميات مختلفة تتصل بالأطر مثل إطار كرانفليد Cranfield Paradigm أو الإطار الفيزيائى Physical Paradigm أو الإطار النموذجى أو المثالى Archetypal ثم الإطار المعرفى Cognitive، ويذهب البعض إلى أنه على الرغم من أن معظم بحوث استرجاع المعلومات تتعامل داخل الإطارين الفيزيائى والمعرفى ، إلا أن أساس الإطار القوى الذى يعكس العلم الحقيقى Normal Science لا يتحقق مع أى واحد منهما ، فما الذى نعنيه بمفهوم الإطار وما دوره فى علم المعلومات وفى بحوث استرجاع المعلومات .

١ - مفهوم الإطار ودوره في علم المعلومات :

لقد وضع الفيلسوف كوهن Kuhn مصطلح الإطار Paradigm كمفهوم لعالم تطور البحث

العلمى ضمن كتابه عن تركيب الثورات العلمية (Kuhn, T. S. 1962) ، ولم يكن بيان كوهن عن الإطار بياناً مبهراً أو واضحاً ، بل كان يحمل بعض الغموض خصوصاً بالنسبة لعلاقته بطبيعة ودور الأطر ، وعلى سبيل المثال لا الحصر فقد وضعت ماسترمان Masterman عدد (٢١) معنى مختلفاً للإطار كما استخدمه كوهن أى ليدل على المعانى التالية :

(أ) إنجاز علمى عالمى معترف به (ب) فلسفة قضايا (جـ) عمل أساسى كلاسيكى فى مجال معين (د) تأمل ميتافيزيقى ناجح (هـ) تمثيل معيارى (و) مبدأ تنظيمى يحكم الإدراك (ى) طريقة جديدة لرؤية الأشياء ... إلخ ، كما لاحظت ماسترمان إمكانية تقسيم المعانى Senses إلى الأقسام العريضة التالية : (أ) الميتافيزيقية (ب) الاجتماعية (جـ) الشىء المصنوع Artifact ويضم المعنى الميتافيزيقى الإطار الذى يرى كمجموعة من الاعتقادات أو كمعيار أو كطريقة جديدة لرؤية الأشياء أو مبدأ تنظيمى لجانب من الحقيقة ، أما المعنى الاجتماعى فيعتبر الإطار كإنجاز علمى محسوس معترف به عالمياً ، أما المعنى الثالث فهو أكثرها تماسكاً وثباتاً Concrete ويدل الإطار هنا على العمل الكلاسيكى أو الكتاب المرجعى الأساسى Text Book (Masterman, M, 1970) .

وقد تقبل كوهن النقد الذى أثير حول أفكاره الأصيلة ، كما ينبغى الإشارة أيضاً إلي صياغاته الأساسية عن تحول الحقل إلى النضج ، أى إلى ما يسميه كوهن بالعلم القياسى أو المضبوط أو الطبيعى أو الحقيقى Normal Science ، وأن هذا العلم يتميز بوجود إطار واحد للمجال يقوم على توحيد الباحثين حوله ، واعتبر كوهن أن العلوم الطبيعية (كالكيمياء والفيزياء) تعتبر أطراً ناضجة بينما تعتبر العلوم السلوكية والاجتماعية وكذلك علم المعلومات من الحالات قبل الإطارية Preparadig-matic أى أن هناك طريقين أساسيين للتعرف على طبيعة الإطار ، وأول هذه الطرق رؤية الإطار بمعنى محدود ضيق كتمثيل حلول المشكلات نموذجياً Representing Model Problem Solutions أما المدخل الثانى لمعنى الإطار فهو توسيع تفسير معنى الإطار طبقاً لصياغة كوهن الأصلية لما يسمى بالمصنوفة الموضوعية Disciplinary Matrix أى محاولة اكتشاف الوضع الإطارى (ذى الأقسام العريضة الميتافيزيقية والاجتماعية والمحسوسة وضمن المدخلين النموذجى والمعرفى) ، وهذا التفسير العريض هو الذى ستتبعه الباحثة فى هذه الدراسة .

٢ - الأطر النموذجية والمعرفية في بحوث استرجاع المعلومات :

Archetypal and Cognitive Paradigms in Information Retrieval Research

على الرغم من التاريخ النسبى القصير لتطور بحوث استرجاع المعلومات كمجال له هويته ضمن علم المعلومات ، إلا أنه مجال يتميز بالثراء فى الإنتاج الفكرى لغرض التحليل .. ولعل هذا التاريخ يبدأ من الاختبارات الأمبيريقية التى أجريت خلال الخمسينيات والستينيات والمتمثلة فى اختبارات كرانفيلد للغات ونظم التكشيف ويطلق على المدخل المتبع بالمدخل النموذجى أو المثالى Archetype أما

المدخل الثانى والذى ينبغى من الاهتمام بالمداخل المعرفية فى الاسترجاع فيمكن أن نعود به إلى منتصف السبعينيات ، وقد رأى البعض أن المدخل النموذجى المتصل بمقارنة نظم ولغات التكشيف يمثل أقرب الجوانب والدراسات للإنجازات العلمية المحسوسة والتى برزت فى علم المعلومات لأول مرة .

كما أن الإطار الميتافيزيقى Metaphysical المرتبط بالمدخل النموذجى فقد تدعم وجوده ، ذلك لأن نظام استرجاع المعلومات يعتبر كنظام تشغيلى Operational أكثر منه نظاماً فيزيائياً Physical System وأن البيئة الاختبارية شبيهة بالمختبر ، وهناك تمييز واضح بين الاختبار التشغيلى Opera-tional والاختبار التجريبى ؛ حيث يتطلب الأخير إنشاء بيئة اصطناعية ، هذا واتجاه البحوث المعتمدة على الاختبارات تعد معياراً للتمييز بين المشكلات الحقيقية والمشكلات الميتافيزيقية ، كما أن الاختبارات تمثل ثورة فى طريقة وضع نظم استرجاع المعلومات ، فقبل اختبارات كرانفيلد كان نظامى التصنيف والتكشيف يعتبران – بطريقة أو أخرى – نظم شبه ميتافيزيقية Quasi-Metaphysical .

أما المداخل الحصرية التقليدية للتصنيف فقد مثلت المعرفة بصفة عامة .. أى أن هناك القليل من التبرير المعرفى Epistemological Justification للاعتقاد بأن نظم التصنيف الناتجة تمثل خططاً للمعرفة Schemes of Knowledge كما أن المدخل النموذجى أو المثالى Archetypal Ap-proach له أصوله فى اختبارات أدوات ونظم التكشيف وذلك منذ أواخر الخمسينيات وبداية الستينيات ، واعتبرت قياسات كرانفيلد الإطار الأنطولوجى والمنهجى Ontological and Me-thodological نشأ داخله هذا التخصص الفرعى . وكانت نتائج هذا القياس الرئيسية أقرب الإنجازات العلمية التى حققها العلم لأول مرة (Jones, Spank 1981) .

هذا وتعتبر اختبارات كرنفيلد نماذج كلاسيكية للبحث فى حقل المكتبات والمعلومات ، وتم قبولها كمقدمة منهجية للممارسة المستقبلية . وقد أثرت وأثرت فى اختبارات كرانفيلد على الكتب الأكاديمية فى حقل المعلومات ومن أمثلة هذه الكتب :

- **نظم استرجاع المعلومات** : خصائصها واختباراتها وتقييمها (١٩٧٩) (لانكستر) .
- **مقدمة لاسترجاع المعلومات الحديثة** (١٩٨٣) (سالتون وماكجيل Salton & McGill) .
- **استرجاع المعلومات** (١٩٧٩) (فان ريجسبرجن Rijsbergen) .
- **مقدمة فى استرجاع المعلومات الحديثة** (١٩٩٩) (شاودورى Chowdhury) .

وليس من المعقول التفكير فى مثل هذه الكتب مجردة ومسلوبة من إطارها الأمبيريقى والمفهومى والذى يعتمد فى النهاية على تجارب كرانفيلد على مدى أكثر من عشرين عاماً وعلى الرغم من وجود نقد لهذه الأدوات ووجود الإضافات والتعديلات التى تمت فى البحوث التالية لتجارب كرانفيلد ، إلا أن تجارب كرانفيلد تظل أساساً مرناً وعملياً يدعم معظم الأعمال الأمبيريقية التى جاءت من بعدها .

كما تأتينا الأدلة الإضافية للهوية البنائية عن طريق المكون القوى المتماثل بين كل من البحوث القديمة والحديثة ، خصوصاً بالنسبة للمراجع العديدة المتصلة بالمفاهيم والنظم الفيزيائية والميكانيكية ، ذلك لأن نظم التكشيف تعتبر ممثلة لخليط أو مزيج من الأدوات الكشفية Indexing Devices فضلاً عن التعبيرات التى نتداولها كفراغ الوثيقة Document Space ، وتقارن التطورات فى نظم استرجاع المعلومات بالتطورات فى السيارة حيث ارتبطت بيئة اختبارات كرانفيلد بنفق الريح Wind Tunnel (Cleverdon, C. W., 1971) موصولة بالميكانيكا الكمية وتترك حتى القراءة السطحية للإنتاج الفكرى شكاً قليلاً بالنسبة للطبيعة الفيزيائية للتشابه الذى يكمن وراء البحث عند اتباع هذا المدخل .

وتدعم المقارنة المباشرة بين خواص مفهوم كوهن (Kuhn's Concept) ومفهوم القوالب التخصصية Disciplinary Matrix الخصائص الإطارية الواضحة لهذا المدخل ، فعنصر القوالب التخصصية الذى يمكن تحديده واضح فى التعميم الرمزى الذى تمثله جوانب الصلاحية Relevance المعتمدة على نسب الأداء (وهي الاستدعاء والدقة والشوشرة ..) وعلى صيغة العلاقة العكسية Inverse Relationship بين الاستدعاء والدقة .. ويقع الإطار الميتافيزيقى فى علاقة التلازم المشتركة للمعتقدات ، بأن نظم استرجاع المعلومات تتكون من عناصر أو أدوات مختلفة ، وأن دور نظام استرجاع المعلومات هو استرجاع جميع الوثائق المتعلقة بالمستفيد فضلاً عن الالتزام بالافتراضات المعرفية والسلوكية لنموذج استرجاع المعلومات .

هذا ويبدأ المدخل المعرفى من افتراض مفاده أن التفاعل الفعلى بين المستفيد ونظام المعلومات يتم من خلال تفاعل أكثر تعقيداً من ذلك المفترض فى المدخل النموذجى Archetypal ، وأنه ذاتى وديناميكى خلال التفاعل بين الباحث والنظام هذا والصورة المجازية Mataphore وراء المدخل المعرفى هو تخلى النظام بشىء مثالى للعالم أو أنه نظام تمثيلى للمفاهيم أو الفئات التى تتوسط التفاعل بين المستفيد والنظام (De May, M., 1982) ، وإذا كانت وجهة النظر المعرفية لعلم المعلومات تعود فى كثير منها إلى الباحث دماى De May فإن الشخص المفتاحى الرائد فى وجهة النظر المعرفية لعلم المعلومات تعود إلى العالم بروكس (Brookes, B. C., 1980) كأستاذ وإلي العالم بيلكن كطالب لبروكس (Belkin, J. 1980) حيث قام بيلكن بإعادة صياغة المفاهيم للتركيز المباشر على مشكلات استرجاع المعلومات .

وما تجدر الإشارة إليه أن مفاهيم المدخل المعرفى لبحوث استرجاع المعلومات يعتبر أكثر حداثة فى أصوله من المدخل النموذجى Archetypal وبالتالى فالمدخل المعرفى أقل تطوراً فى صفاته الإطارية Paradigmatic Character ، ومع ذلك فميزته الإطارية الأقوى هى فى الإطار الميتافيزيقى أى فى الفلسفة أو مجموعة المعتقدات المتصلة بطبيعة التفاعل الاسترجاعى .

ونخلص من هذا الحوار والمناقشة إلى أن نطاق بحوث استرجاع المعلومات يتكون لدى فيكرى من المعرفة العامة Public Knowledge (مثل الوثائق التى تضم جميع أشكال المعرفة المسجلة أو

المعلومات) وتمثيلات لهذه المعرفة المسجلة والتى يطلق عليها فيكرى تركيبات المعرفة الوسيطة Intermediate Knowledge Structures وهذه تتخذ أشكال التصانيف والكشافات والمستخلصات بالإضافة إلى تركيبات المعرفة الشخصية ذات العلاقة بالمعرفة السائدة واحتياجات المستفيدين (Vickery, B. C., 1992) .

ويعكس هذا التقسيم Typology خواص فئات العالم الثالث التى تعود إلى فكر بوبر (Popper, K. R., 1972) حيث يحتوى العالم الأول على عالم الأشياء المادية Physical Pbjects، أما العالم الثانى فهو عن العقول ، أما العالم الثالث فيضم التراث الثقافى . وتعتبر كلاً من المعرفة العامة والمعرفة الوسيطة وهى التى توازى مفاهيم العالم الثالث جزءاً من وسائط العالم الأول كما أن المعرفة الشخصية توازى العالم الثانى .

وأكد فيكرى على أن البحوث التقليدية لكرانفيلد تُظهر الاتجاه نحو التركيز على تركيبات المعرفة الوسيطة ، وليس على المعرفة العامة أو الشخصية ، وإن كانت هناك بعض الآراء الحديثة التى تشير إلى أن هذا الاتجاه قد أصبح أقل تأثيراً من ذى قبل ، خصوصاً وقد سمحت وثائق النص الكامل بإمتداد اتجاه مدخل كرانفيلد إلى النصوص الكاملة ، وبالتالى قدمت لنا فرصة للتحول بعيداً عن مجال البحث المتصل بتركيبات المعرفة الوسيطة وقريباً من المعرفة العامة ، طالما كانت هذه المعرفة العامة موازية لتمثيلها فى الوثائق أو غيرها من أشكال الوسائط المسجلة . ويعتبر فروهمان (Frohman, B., 1990) أن هذه معادلة لا يستطيع أحد الدفاع عنها Untenable .. ولكن هذه الاتجاهات جميعها لا تظهر نفس مستوى الالتزام نحو فهم ونمذجة تركيبات المعرفة الشخصية والمعرفة العامة وهى التبرير العقلانى للبحث فى المدخل المعرفى (Ellis, D., 1998: 14) .

وبناء على ذلك فإذا كان هناك تقارب بين المدخلين فى تصميم البحوث الفعلية ، فإن ذلك يتم من منظورين مختلفين تماماً ، أحدهما يميل إلى التركيز على التمثيلات السطحية للمعرفة المسجلة فى الوسائط المادية ، والآخر يعرض الاتجاه المعاكس ويحاول تحديد وتمثيل التركيب الأكثر عمقاً ، المشمول فى كل من تركيبات المعرفة العامة والمعرفة الشخصية .. وعلى الرغم من أن البعض يرى أن جميع البحوث الاجتماعية يمكن أن تصنف إما كبحوث موضوعية Objectivist تماماً أو بحوث ذاتية Subjectivist تماماً ، وهذا ما يظهر بوضوح أيضاً فى بحوث استرجاع المعلومات ، إلا أن هذا الاتجاه هو اتجاه خاطئ ويعكس عدم الفهم البحثى السليم .

فالمدخل الموضوعى Objectivist يرتبط بشدة بالانطولوجيا والابستومولوجيا ومنهجية اختبارات كرانفيد كما أن المدخل الذاتى Subjectivist يمكن أن يرى فى ارتباطه الشديد أيضاً بالأنطولوجيا والابستومولوجيا ومنهجية المدخل المعرفى (Ellis, C. 1998) .

وعلى الرغم مرور عدة عقود من البحوث فى دراسات كرانفيلد ، ومرور أكثر من عقد ونصف على البحوث فى الاتجاه المعرفى ، فالتطور نحو وجود جسد متماسك من المعرفة فى استرجاع

المعلومات مازال مخيباً للآمال Disappointing وذلك على الرغم من وجود ما يمكن أن يطلق عليه تحول إطارى من دراسات كرانفيلد إلى المدخل المعرفى . هذا ويكشف تحليل الإنتاج الفكرى لبحوث استرجاع المعلومات أن الباحثين الذين يتبعون مدخل كرانفيلد يستخدمون مجازات مادية Physical Metaphores وإشارات Allusions بنفس الطريقة التى يتبعها الباحثون فى المدخل المعرفى Cognitive ، وما يمكن ملاحظته أن المدخل الفيزيائى Physical Approach يتميز بجسد واضح من البحث الأمبيريقى فى اختبارات كرانفيلد ، ولكن دون وجود صورة واضحة غير مختلف عليها Unambiguous لطبيعة الإطار المتبع . أى أن المدخلين المعرفى والفيزيائى يفتقدان الهوية الإطارية والتى يمكن أن تشكل أساس علم متطور بصورة كاملة .

رابعاً : مشكلات القياس ودورها فـــي أطـــر بحوث استرجاع المعلومات :

تشير الباحثة فى البداية إلى قلة دراسات القياسات فى علم المعلومات بصفة عامة وفى استرجاع المعلومات بصفة خاصة فى الإنتاج الفكرى باللغة العربية ، ويمكن أن تعود مشكلات القياس هذه إلى اختبارات الاسترجاع الأولى والتى أخذت أشكالاً ثلاثة مع تطور المجال وهى : (أ) مشكلات القياس فى المدخل الأصلى (أو النموذجى أو الفيزيائى Archetypal) (ب) مشكلات القياس فى المدخل الاحتمالى Probabilistic (جـ) مشكلات القياس فى نظام الخبير لاسترجاع المعلومات IR-ES Information Retrieval Expert System .

وظلت مشكلات القياس فى استرجاع المعلومات قائمة دون حل جذرى وذلك للأسباب الثلاثة التالية : طبيعة المادة الموضوعية فى المجال ، طبيعة الحكم على الصلاحية ، ثم طبيعة المكون المعرفى Cognition and Knowledge .

ويمكن الإشارة لبعض المقتطفات من دراسة دافيد أليس (Ellis, D., 1996) عن مشكلات القياس فى المداخل الأصلية والاحتمالية ونظام الخبير .

(١) مشكلات القياس فـي المدخـل الأصلـي (او النموذجى او الفيزيائى) Afrchetypal :

لقد كان الاستخدام الأول لمصطلح الصلاحية Relevance كمعيار أداء لنظام استرجاع المعلومات، وكان معيار الفاعلية Effectiveness المستخدم هو صلاحية الوثائق لأسئلة حقيقية . ولكن هذا الاختبار فشل بالنسبة لقضية الحكم على الصلاحية .. ولم تستطع الجماعتان الموكلتان بالاختبار الوصول إلى اتفاق حول أى الوثائق تكون أكثر صلاحية لأي الأسئلة وكان ذلك بالنسبة لاختبارات كرانفيلد الأولى ، حيث كان لكل جماعة من الجماعتين تفسيرها الخاص للسؤال ، وبالتالى رؤيتها لصلاحية الوثائق المتعلقة ، وفى اختبارات كرانفيلد الثانية كان هناك نوعان من الصلاحية هما صلاحية المستفيد User Relevance والصلاحية المشروحة Stated Relevance حيث تتقرر صلاحية المستفيد بالسائل نفسه ، أما الصلاحية المشروحة فتتقرر بواسطة أى شخص له معرفة معقولة بالحقل العلمى ، ومن بين التجارب التى تمت على اختبارات كرانفيلد تبين

وجود عدد ضخم من العوامل التى أثرت على انتظام أحكام الصلاحية ، وهذه العوامل تشمل نوع الوثيقة وموضوعها ومستوى الصعوبة وأسلوب الوثيقة ودرجة التركيز ، والزمن المتاح للقائم بالحكم ونظام ترتيب المواد .. إلخ مما يؤدى إلى إنخفاض أو ارتفاع درجات الصلاحية .

ومع ذلك وبعد دراسات وتجارب عديدة فقد تبين أن عدم مقدرة الإنسان على تحديد احتياجاته الفعلية بشكل مبسط ، سيؤدى إلى سحب السجادة من تحت أقدام العديد من دراسات التقييم لنظم استرجاع المعلومات .

كما انتهى بعض الباحثين إلى أن نتائج الاسترجاع Retrieval اعتماداً على الاستدعاء والتدقيق Recal & Precision الذى تم الحصول عليها لمختلف مجموعات الصلاحية المختلفة متماثلة Identical على الرغم من أن الاتفاق العام بين تقييمات الصلاحية تصل إلى حوالى ٣٠ ٪ فقط .

The retrieval results in terms of recall & precision obtained for the various relevance sets are substantially identical even though the overall agreement among the relevance assessments is only about 30 % (lesk & salton, 1968)

(٢) مشكلات القياس في المدخل الاحتمالي Probabilistic

يمكن أن يرى المدخل الاحتمالى فى بحوث استرجاع المعلومات كتمثيل للتطور الخطى Linear Evolution للمدخل الأصلى الفيزيائى Archety-pal وذلك من حيث استخدام المدخل الاحتمالى لنفس الأدوات المفهومية والمنهجية ، ولكن داخل إطار أكثر تنظيراً اعتماداً على فكرة احتمالية الصلاحية ، واعتماداً على أساليب أكثر تحسيباً واحصائية فى استخدام اختبارات المجموعات.

هذا وتختلف مشكلات القياس فى استرجاع المعلومات بصفة عامة عن مثيلاتها فى العلوم الطبيعية (كالكيمياء والفيزياء) من حيث وجود الترتيب الإمبيريقى للكميات المطلوب قياسها فى العلوم الطبيعية . أى أن مقاييس استرجاع المعلومات لا يتوفر فيها الترتيب الأمبيريقى ، وبالتالى فأى مقياس لها هو مقياس اصطناعى بالضرورة ، كما أن مقياس التدقيق والاستدعاء يختلف من وجوه عديدة عن مقياس المتغيرات فى العلوم الطبيعية .. فعلى الرغم من إمكانية وضع أفكار موضوعية للصلاحية ، فالعديد من الباحثين يعتقدون بأن الصلاحية هى بالضرورة مجال ذاتى Subjective أى أن السؤال الواحد الموضوع بواسطة مستفيدين مختلفين ، سيكون له وثائق مختلفة توصف بالصلاحية .

وعلى كل حال فالحيرة والمشكلة فى الاتجاه أو المدخل الاحتمالى هى فى زيادة التحسيب والإحصاء عن طريق اختبارات المجموعات .

(٣) مشكلات القياس في مدخل نظام الخبرة لاسترجاع المعلومات :

تكمن المشكلة هنا فى أن الأساليب والافتراضات المستوردة من بحوث النظم الخبيرة والذكاء الاصطناعى ووجهات النظر المعرفية ، لا تتلاءم مع المفاهيم والأساليب السابقة المرتبطة باستخدام أحكام الصلاحية كمعيار للقياس ، كما

تكمن المشكلة فى كيفية قياس المتغيرات فى تركيبات المعرفة وذلك للاستجابة لاسترجاع المعلومات ، وإذا كانت الصلاحية تعتبر كمقياس فهناك انقسام Schism بين الافتراضات النظرية لتصميم البحث وشكل القياس المستخدم فى التقييم.

وخلاصة هذه المشكلات والحيرة المتصلة بالقياس فى بحوث استرجاع المعلومات ، هى أنه على الرغم من التزامها المبكر بالطريقة العلمية فى القياس ، فلم تصل هذه البحوث إلى أى إطار كامل Articulated Paradigm لتوجيه البحوث (Ellis, 1992, 1994) والمشكلات التى كانت موجودة عند تأسيس المجال مازالت تواجه الباحثين المعاصرين عند تحليلهم لمشروعات حديثة مثل تريك والتى بدأت عام ١٩٩١ فى الولايات المتحدة (TREC = Text Retrieval Confernce) .

نبذة عن مؤتمرات استرجاع النص (TREC) :

لقد كانت المشكلة الرئيسية أمام الباحثين فى تقييم الاسترجاع هو عدم وجود مجموعات ضخمة لمضاهاة المواقف الفعلية ، وقد أنشئت مشروعات تريك TREC لمواجهة هذا النقص أي أن أهداف تجارب تريك TREC تتركز فيما يلى : (أ) زيادة البحوث فى استرجاع المعلومات مع مجموعات اختبار ضخمة (ب) زيادة الاتصال بين الأكاديميين العاملين فى البحث مع رجال الصناعة وغيرهم (جـ) زيادة التكنولوجيا المستخدمة فى كل من البحث والمنتجات (د) تحسين أساليب التقييم :

وهناك استراتيجيات عديدة تعاملت مع بحوث تريك TREC منها :

(١) الاسترجاع البولينى Boolean (٢) التكشيف الإحصائى والاحتمالى واستراتيجية وزن المصطلحات (٣) استرجاع الجمل (٤) تجميع نتائج أكثر من بحث واحد فى المجال (٥) التكشيف اعتماداً على اللغة الطبيعية والتكشيف الإحصائى (٥) البحث المعتمد على الجذور Stemming .

وواضح أن بحوث تريك TREC قد أعادت الحيوية إلى المجال وأبرزت أهمية استرجاع النصوص للباحثين الأكاديميين ورجال الصناعة والحكومة وغيرهم .

ويذهب الباحث شودورى (Chowdhury, G. 1999 : 422) إلى أن تقييم نظم الاسترجاع فى بيئة مختبرية لها تاريخ يزيد عن الأربعين عاماً ، وإذا كانت الاختبارات الأولى قام بها كرانفيلد إلا أن البرامج التجريبية الحديثة يمثلها تريك TREC وهذه وتلك تعكس إطاراً قوياً Powerful Paradigm لمدى واسع من مجالات استرجاع المعلومات .

لقد استمرت مشكلة القياسات دون إمكانية السيطرة عليها نظراً لطبيعة المادة الموضوعية للمجال، وطبيعة الحكم على الصلاحية ، وطبيعة المكون المعرفى Cognition and Knowledge ، وبالتالى فيرى العديد من الدارسين أن مشكلات القياس فى استرجاع المعلومات شبيهة بمشكلات علم النفس وليس بمشكلات الفيزياء .

خامساً : تقييم محركات البحث علي الوب ودورها في الاسترجاع :

مقدمة :

تُستخدم الإنترنت فى الوقت الحاضر فى جميع أوجه تداول المعلومات ، فالمكتبات مثلاً تقدم فهرسها العام (الأوباك) OPAC على الإنترنت وبائعوا قواعد البيانات Vendors على الخط المباشر مثل الديالوج والأقراص المكتنزة CD-ROM مثل سيلفر بلاتر هذه وغيرها تجعل قاعدة بياناتها متاحة على الإنترنت .

هذا وتعتبر مصادر المعلومات على الإنترنت ملفات تحتوى على المعلومات فى الشكل الرقمى (النصوص والصور والأصوات والفيديو ..) فضلاً عن البرامج .. كما أصبحت محركات البحث أكثر الأدوات أهمية لاسترجاع المعلومات على الوب وأكثر المجالات أهمية فى البحث بعد وصول المعلومات على الوب ، ومحرك البحث على الإنترنت هى خدمة استرجاع تحتوى على قاعدة بيانات تصف أساساً المصادر المتاحة على الوب ، أما صفحات الوب فتحتوى على معلومات الوسائط المتعددة Multimedia المصممة باستخدام لغة علامات النص الفائق HTMI ويمكن عنونتها من خلال محدد المصادر المنتظم URL (Uniform Resource Locator) .

ومع تضخم حجم الوثائق وأنواعها على الوب ، أصبحت بيئة الوب غير محكومة Uncontained أى أن الباحث لا يستطيع التأكد من حجم الوثائق التى يجب استرجاعها أو المتاحة فى إجمالها ، ويذهب البعض إلى أن الاستدعاء Recall أصبح مستحيلاً حسابه فى هذه البيئة السريعة التغيير ، وبالتالى فإن طريقة كرانفيلد التقليدية فى التقييم أصبحت غير ممكنة .

فعلى الرغم من أن محركات البحث تقوم ببحث حجم ضخم من المعلومات بسرعة هائلة إلا أنها عرضة للنقد من حيث وقت الاستجابة البطيء واسترجاع تسجيلات مكررة والفشل فى استرجاع مواد متعلقة وصالحة وذلك مع استرجاع كمية ضخمة من المواد غير المتعلقة .. أى أن المواد الجيدة يصعب الحصول عليها (Good Stuff (Lowley, S., March 2002) . is Hard to Find)

انواع محركات البحث :

يمكن تقسيمها إلى الفئات الأربعة التالية :

الروبوت Robots – الأدلة Directories – محركات البحث الوسيطة Meta Search Engines وأدوات البرامج وتجمع بعض المحركات بين صفات أكثر من فئة (Lowley, S., March, 2000) .

(أ) محركات البحث المعتمدة علي الروبوت :

تولد برامج الزواحف Worm or Crawler قواعد بيانات عن طريق روبوت الوب ويتم الاسترجاع من المواقع على الوب باستخدام البرتوكولات المعيارية ومن بين محركات البحث الهامة : التافيزتا / اكسيت / ليكوس / هوتبوت Hotbot / انفوسيك .

(ب) محركات البحث المعتمدة علي الأدلة :

وتعرف هذه المحركات بأنها مجموعات موضوعية أو بوابات الموضوع Subject Gateways وهى مجموعات لربط الـ URL المتعلقة ويتم صيانتها بواسطة الخبراء وبعض طرق التكشيف الآلى ، وتقدم روابط الهيبرتكست المستفيد إلى الـ URL الأكثر ملاءمة ومن أمثلة المحركات المعتمدة على الأدلة يا هو Yahoo ، دليل المملكة المتحدة UK Directory .

(جـ) محركات البحث الوسيطة

Meta Search Engines

وتستخدم هذه المحركات الوسيطة قواعد البيانات التى يتم توليدها بالمحركات الأخرى ، ويتقبل محرك البحث الوسيط سؤال المستفيد ثم إرساله إلى محركات بحث متعددة بالتوازى ومن بين المحركات الأكثر شعبية ميتا كرولار Meta Crawler .

(د) ادوات البرامج Software Tools :

وتصنف هذه الأدوات أحياناً كمصاحبة للتصفح Browsing Companions وهي شبيهة بمحركات البحث الوسيطة ولكنها تحتاج إلى تركيبها على محطة عمل المستفيد ، ومن أمثلتها ويبزلتيث Wevbsleuth وكوبرنك Copernic 98 وميتاهارى Meta Hari .

بعض النتائج :

هناك العديد من الجهود والدراسات للقيام بمقارنات عن مدى فاعلية محركات البحث ويعتبر البعض أن منهجية التقييم لمحركات بحث الوب يمكن أن تشمل ما يلى :

(أ) مقارنة كشافات الوب بما فى ذلك التغطية والحداثة وأجزاء صفحات الوب التى يتم تكشيفها (مثلاً العناوين + السطور الأولى ... أو جميع صفحات الوب) .

(ب) إمكانيات البحث بما فى ذلك منطقية بوليان Boolean ، البحث عن الجمل ، البتر ..

(جـ) الأداء الاسترجاعى المعتمد علـــى الدقة Precision والاستدعاء Recall .

(د) اختيار المخرجات Output Option اعتماداً على التحليل النوعى والكمى .

(هـ) مجهود المستفيد اعتماد على تحليل التوثيق والتفاعل Interface .

وأخيراً يمكن الإشارة لجانبين فى تطور محركات البحث أولهما : تجارب تريك الحديثة (TREC-4) والتى تتناول ٣,٢ جيجابيت من النصوص قد وصل إلى أسلوب جديد لتقدير الاستدعاء Recall ومع ذلك فإن ما نحصل عليه هنا هو استدعاء نسبى وليس استدعاء مطلق أي أن هذا القياس لا يأخذ فى اعتباره المواد الصالحة Relevant Items التى يفشل الباحث فى استرجاعها .

وثانيهما هو محركات البحث الذكية : وهذه بدأ التفكير فيها عام ١٩٩٤ حيث أشار تقرير مجلس البحوث الوطنى الأمريكى إلى قضية رئيسية للإنترنت ، وهى كيفية الملاحة وترشيح النص ، وأشار التقرير إلى أن الحل الأساسى يأتى من بحوث التفاعل مع المستفيد User Interfaces وليس من مجرد طرق البحث .. وان المكون الذكائى هو البرنامج الذى يمكنه مساعدة الناس والعمل بالنيابة

عنهم ، وإذا كان هناك على الأقل خمسين بائع (بما فيهم IBM) يقدمون البرامج والخدمات ، فإن واحداً من هذه الفئات هى محركات البحث المصممة لإلتقاط المعلومات المتاحة من مختلف المواقع على الوب . وتتعلم هذه البرامج من خبرة المستفيد ويمكن أن تعمل بالنيابة عنه ومثل هذه البرامج الخاصة بالوكيل Agent Programs تقوم بعدة وظائف مثل البحث عن المفاهيم Concepts وليس الكلمات المفتاحية Keywords ، ترشيح ووضع فئات وتحرير المعلومات .. أي أن وكلاء البحث الأذكياء Intelligent Search Agents يتم تعميمها فى الوقت الحاضر للقيام بعدة عمليات فى بيئة المعلومات الرقمية ، وحتى يمكن حل المشكلات القائمة فيجب على المهنيين فى المعلومات – بخبرتهم الثرية فى التنظيم والبحث والتعرف على احتياجات المستفيد – العمل مع المهنيين فى البرمجة Software Professionals فى تنظيم ما يمكن تسميته بفوضى الإنترنت Anarchy of Internet . (Chowdhury, G., 1999)

سادساً : استرجاع المعلومات الذكية وحل بعض المشكلات التي تعاني من البحث البوليني :

Intelligent Information Retrieval and Boolean Logic

مقدمة :

يعرض الباحثون كورتيز Cortez وبارك Park وكيم Kim (Cortez, E. M., 1995) تجربتهم فى استرجاع المعلومات الذكية عن طريق التطبيق المهجن Hybrid للتعلم الاستقرائى والشبكات العصبية Inductive Learnig Method and Neural Network ، وواضح استخدام مجالات علمية جديدة لمحاولة حل مشكلات استرجاع المعلومات المعاصرة والمزمنة .

هذا ويتركز الهدف الرئيسى لنظام استرجاع المعلومات (IR) فى الاستجابة للأسئلة الخاصة بالمعلومات فى أى موضوع محدد ، وذلك عن طريق استرجاع مجموعة من الوثائق المتعلقة ببعضها، وتعتمد معظم نظم استرجاع المعلومات التقليدية على المنطق البولينى ، حيث تتيح لنا عملية الاسترجاع البولينى ملفات وثائق لمجموعات فرعية بناء على مضاهاة أو عدم مضاهاة بين مصطلحات السؤال ومصطلحات تكشيف الوثيقة ، وبمعنى آخر فتعتمد نظم استرجاع المعلومات على نموذج الاسترجاع البولينى الذى يفترض أن السؤال والوثائق يتم تمثيلها بدقة بواسطة مجموعة من الكلمات المفتاحية فى اللغات الطبيعية . (Belking et al, 1982)

ونتيجة لذلك فينتج النظام مجموعة من الوثائق القريبة من بعضها والتى تضاهى بدقة السؤال ، فى الوقت الذى ترفض فيه كل الوثائق الأخرى القريبة جزئياً أو التى لا تنطبق عليها المضاهاة non-matching documents ونظام الاسترجاع النموذجى يعتمد على المنطق البولينى ويكشف مشكلات داخلية مرتبطة بفاعلية الاسترجاع .

هذا وتعانى معظم نظم الاسترجاع التقليدية التى تعتمد على المنطق البولينى من مشكلتين أساسيتين هما (أ) تمثيل الأسئلة بمصطلحات غير

دقيقة أو غير محددة دلالياً مما يؤدى إلى إنخفاض فاعلية الاسترجاع (ب) التكشيف غير المنتظم أو غير الكامل خصوصاً مع إضافة وثائق للمجموعات على فترة طويلة من الزمن بواسطة باحثين مختلفين، وإذا كان العديد من الباحثين قد دللوا على إمكانية حل مشاكل الأسئلة غير المحددة عن طريق الشبكات العصبية Neural Networks ، إلا أن مشكلة التكشيف غير المنتظم مازالت فى حاجة إلى الحل . وتقدم لنا هذه الدراسة منهجية تهجين يتم بواسطتها تكامل أسلوب التعلم الاستقرائى مع الشبكة العصبية (نموذج الربط Connectionist Model) من أجل حل المشكلتين المتصلتين بالتكشيف غير المنتظم ومشكلة الأسئلة غير الكاملة أو غير الدقيقة ، ونظراً لقدرة أسلوب التعلم الاستقرائى على تحديد أكثر المصطلحات الكشفية ذات الدلالة للوثيقة بمستوياتها المختلفة فى علاقتها الدلالية Semantic Significance فهى تزودنا بالحل الممكن لمشكلة التكشيف غير المنتظم ، وتقدم لنا دراسة كوريتز Cortez وزملاؤه المرحلة الأولى من البحث الذى يبين لنا كيفية إضافة Augement الشبكة العصبية لأسلوب التعلم الاستقرائى للوصول إلي أداء استرجاعى فعال للمعلومات ، وذلك فى المجالات التى تتطلب استنتاجاً وتبريراً مرناً للأسئلة غير الكاملة أو مشكلات التكشيف غير المنتظم .

وتتضمن الصفات الخاصة بنظام استرجاع المعلومات الذكى المقترح جانبين هما : (أ) يعوضنا هذا النظام المقترح عن عدم الدقة وعدم الاكتمال فى السؤال ، ذلك لأن نظم الاسترجاع التقليدية تتطلب مجموعة كاملة – إلى حد ما – من الكلمات المفتاحية المتعلقة بسؤال المستفيد ، أما فى النظام المقترح فالمستفيدون قادرون على بدء استرجاع الوثائق المتعلق عن طريق استخدام مصطلحات السؤال فقط ، وخلال عملية الاسترجاع فإن مصطلح السؤال النشط Active Query Term (مثلاً المصطلح الشبكات العصبية) يمكن أن يؤدي إلي مصطلحات أخرى متعلقة بالسؤال (أى التعلم) ، هذا ومجموعة الوثائق المسترجعة هى تلك التى ستضاهى Match مصطلحات السؤال وهى المصطلحات النشطة فى البداية ، أو أنها المصطلحات التى ستنشط داخلياً Internally Activated . ومعنى ذلك أن النظام لديه القدرة على التفسير المرن للسؤال ، وتسمح هذه المرونة للنظام فى تعويض الأسئلة غير الدقيقة أو غير الكاملة .

(ب) يمكن للنظام تعويض عدم الانتظام فى تكشيف مجموعات الوثائق ، ذلك لأن النظام المقترح يبنى تركيبة خريطته من بين مصطلحات الوثيقة ، فأسلوب التعلم الاستقرائى له مقدرة تمييز أكثر مصطلحات التكشيف ، أى المصطلحات ذات الدرجات المختلفة للمعانى الدلالية . ومعنى ذلك أن المعنى الذى يصحب الكلمة بواسطة المستفيد ، لا يعكس بالضرورة المعنى المختزن فى هذه الوثائق ، وبالتالى فنظام استرجاع المعلومات الذكى ، يمكن أن يساعد فى السيطرة على عدم الانتظام وعدم الاكتمال فى تكشيف المجموعة .

وفى الملاحظات النهائية لنتائج الدراسة يذهب الباحثون إلى أنه بالرجوع إلى السؤال البحثى للمشروع : هل ستقوم المنهجية المهجنة (التى يتم

بواسطتها التكامل بين أسلوب التعلم الاستقرائى مع الشبكات العصبية) بحل مشكلات استرجاع المعلومات والتى تحتوى التكشيف غير المنتظم والأسئلة غير الكاملة ؟ ويجيب الباحثون القائمون بالمشروع بأنهم يعتقدون بإمكانية الحل بواسطة هذا الأسلوب المقترح .

سابعاً : موقع الانترنت (الوب) فى الخدمة المرجعية الرقمية بالمكتبات الاكاديمية :

(١) تعريف المكتبة الرقمية وبعض قضاياها البحثية :

يعنى مصطلح المكتبة الرقمية أشياء مختلفة للعديد من الناس ، فيعتبرها البعض مجموعة الوثائق فى الشكل الرقمى والتى تسمح للمستفيدين بتوليد واستخدام المعلومات إلكترونياً ، بينما يعتبرها البعض الآخر مكتبة افتراضية Virtual على الإنترنت ، حيث تتاح جميع معلوماتها فى الشكل الرقمى بينما يستخدم البعض مصطلح المكتبة الرقمية فى تبادل مع مصطلح المكتبة الالكترونية وهل هى البنية التحتية المعلوماتية التكنولوجية Technological Infrastructure أم أنها اختزان إدارة المعلومات الرقمية أم أنها هذا كله ؟

وإذا أخذنا المعنى العام لنظام المكتبة الرقمية وأنه عكس نظم المكتبة التقليدية فالمكتبة الرقمية ليس لها بناء محدد فى منطقة معينة ، ذلك لأنها موزعة فى الواقع فى جميع أنحاء العالم ويستطيع المستفيدون الحصول على المعلومات عند الحاجة إليها على شاشات الحاسبات الآلية ، وإلى جانب المعلومات النصية التقليدية فالبيانات المتاحة من خلال نظام المكتبة الرقمية يمكن أن يشمل ما يلى:

- المعلومات غير النصية (مثل الصور والرسومات والفنون وغيرها) .
- بيانات رقمية مستمرة (مثل معلومات الأقمار الصناعية وبيانات الأرصاد الجوية ... إلخ) .
- الصوت الرقمى والصور المرئية المتحركة .
- التمثيل المتعدد الأبعاد للأشكال المختلفة (مثل الهولوجرام Holograms) .
- تكامل هذه البيانات فى تمثيلات جديدة مستمدة من مصادر مختلفة عديدة .

أما بالنسبة لقضايا المكتبة الرقمية البحثية فمازالت هناك مشروعات عديدة فى أنحاء مختلفة من العالم ويمكن أن يؤدى حلها إلى جعل المكتبة الرقمية أداة طبيعية للمستفيد العادى ومن بين هذه القضايا ما يلى :

- كيف يمكن التعرف على المستفيد من المكتبة الرقمية ، وما هى طبيعته واحتياجاته وسلوكه البحثى .. أى أنها طبيعية لا نستطيع التنبؤ بها ؟
- كيف يمكن لنظم المعلومات المختلفة حول العالم أن تتحدث إلى بعضها وتشارك مع بعضها فى المعلومات ؟
- كيف يمكن تطوير واجهة تعامل مبتكرة In-novative Interface تزودنا بالوصول اليسير إلى المجموعات المختلفة وتجعل من اليسير أيضاً على المستفيدين أن يصبحوا هم مزودين للمعلومات عن طريق الإضافة للمجموعات على الخط المباشر ؟

- كيف سيتمكن المستفيدون (أو من ينوب عنهم) من معرفة مكان المعلومات التى يحتاجونها وكيفية الوصول إليها ؟
- كيف يمكن فرز نتائج البحوث واختيارها ودمجها وبثها للمستفيدين ثانية ؟
- كيف يمكن لنظم التوزيع الضخمة أن تتعلم وتنمو ؟
- كيف يمكن ضمان هوية وأمن البيانات ؟
- كيف يمكن ضمان حق المسهمين فى المعلومات ؟
- كيف يمكن تصحيح عمليات سوء استخدام المعلومات .
- كيف يمكن التعرف على أنواع الفشل فى نظم المعلومات وتحديدها ؟

وهناك العديد من الباحثين والهيئات حول العالم الذين يقومون بالبحوث المتصلة بمفاهيم المكتبة الرقمية والعديد منهم يعتقد بصلاحية الأدوات والأساليب التقليدية المستخدمة فى الحفظ والاسترجاع التقليدى مع بعض التعديلات المناسبة ولبيئة المكتبات الرقمية .

فمشروع سكوربيون Scorption هو أحد المشروعات البحثية المركز التحسيب على الخط المباشر (OCLC) .

(http://www.oclc.org/oclc/research/publication/review96/scorpion.htm)

وقد بنى المشروع حول هذا المفهوم ، حيث يحاول الباحث أو الأمين التعرف على إمكانية التكشيف والفهرسة للمصادر الالكترونية ، ونظراً لأن التكشيف الموضوعى يعتبر متفاحياً لاسترجاع المعلومات ، فضلاً عن التصفح ، فيعتبر المحور الأولي لمشروع سكوربيون هو بناء أدوات التعرف الموضوعى الآلى المعتمد على خطط التصنيف مثل تصنيف ديوى العشرى

(http://orc.rsch.oclc.org:6109/bintro.html)

وهناك مصادر عديدة تعكس التطورات فى أنواع مختلفة من مشروعات المكتبة الرقمية مثل مصادر المكتبة الرقمية ومشروعاتها فى مكتبة الكونجرس وصفحة مصادر الإنترنت بالتعاون مع اتحاد المكتبات الرقمية القومية National Digital Library Federation وكذلك مبادرة المكتبة الرقمية Digital Library Initiative (DLI) (http://www.nsf/gov.1) وتركز هذه البحوث على تطوير طرق متقدمة لتجميع واختزان وتنظيم المعلومات فى الأشكال الرقمية فضلاً عن جعل المعلومات متاحة للبحث والاسترجاع والمعالجة من خلال الإنترنت وبطرق صديقة للمستفيد .

وهناك برامج شبيهة موجودة بالمملكة المتحدة مثل برنامج (elib) .

(http://www.ukloln.ac.uk/services/elib)

وهناك عدد من الباحثين المهتمين بالبحث فى طرق فهرسة وتصنيف المواد عبر الإنترنت والشبكة العنكبوتية www (Younger, J. A. 1999) .

(٢) تطور الخدمة المرجعية الرقمية علي الوب :

يقصد بالمكتبات الرقمية على الوب تقديم الإتاحة المباشرة لمصادر وخدمات المعلومات دون

تدخل الوسيط الإنسانى Human Intermediary، وإذا كانت المكتبات الرقمية المعاصرة تركز على الإتاحة أو استرجاع المعلومات الرقمية أساسياً ، فنادراً ما يكون هدفها وممارساتها على جوانب الخدمة والتى كانت تقدم للمستفيدين من المكتبات التقليدية ، هذا والدراسات المعاصرة (Chowdhury, G. C 2002) تعيد تأكيد الحاجة إلى تقديم خدمات المعلومات الشخصية فى الجيل القادم للمكتبات الرقمية وتحاول المشروعات البحثية الجارية التعرف على أفضل الوسائل لتقديم هذه الخدمة عن طريق وسائل عددية منها :

- آليات ترشيح المعلومات المعتمدة على الصفات الشخصية للمستفيدين .
- النماذج التعاونية للمشاركة فى المصادر والتخصصات سواء بالنسبة للمحتوى أو الموظفين .

وقد قام الباحثان تيبنوبير وإنيس (Tenopir, C., Spring 2002) بعمل أربع مسوحات لعينة من المكتبات الجماعية الأمريكية خلال الفترة من (١٩٩١-٢٠٠١) وذلك للتعرف على التغيير الذى حدث بالنسبة للخدمة المرجعية الرقمية وقد اختارت الباحثة هذه الدراسة لشمولها فى المصادر وبعمقها فى الزمن وتعددها فى الخدمات .

أ - **قواعد البيانات المحملة محلية** Locally Loaded وهذه تضم توليفة من قواعد البيانات المرجعية المحملة محلياً على الحاسبات داخل الحرم الجامعى وتلك التى تقدم إتاحة لقواعد البيانات المرجعية على حاسبات المكتبات الأخرى كأعضاء فى الاتحادات Consortiums أى إلى جانب فهارس المكتبة فتقدم إتاحة للمصادر المرجعية فى الإريك ERIC وغيرها .

ب - **قواعد بيانات الأقراص المكتنزة** CD-ROM (سواء التشابكية أو التى تبنى لذاتها) (Networked or Stand Done) وهذه تقدم خدماتها مجاناً وهى تحل رويداً رويداً محل البحث الوسيط على الخط المباشر .

جـ- **بحث الوسيط على الخط المباشر** Intermediary on-line Searching حيث يقوم مهنيون فى المعلومات بالبحث ، وقد انخفضت بنسبة هذه المكتبات الجماعية مع زيادة نسبة المكتبات التى تستخدم الأقراص المكتنزة .

د - **خدمات البحث على الخط المباشر** وهى التى يقوم بها المستفيد النهائى (حيث يقوم المستفيد بالبحث بنظام تجارى على الخط المباشر) ومشكلات وصعوبات الأقراص المكتنزة (كالتشابكية والبطء فى وقت الاستجابة والتقابل Interfaces غير المنتظم) قد جعل البديل على الخط المباشر أكثر جاذبية أحياناً .

هـ - **الإتاحة المباشرة للمستفيد على الوب** حيث تقدم خدمات قواعد البيانات التجارية الروابط إلى مواقع الوب عالية الجودة ، وتساعد على إنشاء مصادر الباحثين الذاتية عن مصادر الوب وتعلم المستفيدين فى كيفية الإفادة من محركات البحث واستخدام الوب فى التعليم المكتبى .

وعلى الرغم من أن هناك أغلبية من المكتبات

البحثية الأكاديمية (ARL) التى تقدم جميع هذه الاختيارات المرجعية ، فالتركيز الواضح على التحول من الخدمات البحثية الوسيطة والأقراص المكتنزة إلى الخدمة الذاتية Self Service سواء على أشكال الـوب Web Versions المعتمدة على الثمن (التجارية) أو الخدمات على الخط المباشر أو موقع الوب الحرة أو المجانية .

أما بالنسبة لكيفية الإجابة على الأسئلة المرجعية ، فعلى الرغم من أن العديد من أقسام المراجع لا تحتفظ بسجل عن المصادر أو الوسائط المستخدمة فى الإجابة علي أسئلة محددة ، إلا أن عدد (٥٨) من المستجيبين من سبعين مكتبة أفادوا بأن معظم الأسئلة المرجعية اعتمدت على الفهرس على الخط المباشر للمكتبة نفسها ، ثم على قواعد البيانات (التجارية) المدفوعة الأجـــر .. وأفـــاد العديد من المستجيبين إلى إنخفاض الاستعانة بالمصادر المطبوعة أمام الوسائل الالكترونية المتعددة المتاحة .

وخلاصة هذه الدراسة المسحية المعمقة أن مصادر المعلومات الرقمية كان لها بدون شك تأثير عميق على الخدمة المرجعية بالمكتبات الأكاديمية والبحثية خلال نهاية القرن العشرين وبداية القرن الحادى والعشرين ، وأن المصادر الإلكترونية ذات أولوية فى بداية البحث ونهايته . فقواعد البيانات التى أصبحت تعتمد على الوب قد غيرت من كل شىء فى هذه الخدمة ، فالطلاب يعرفون كيف يستخدمون الوب وإن كانوا مازالوا يحتاجون إلى المعاونة فى اختيار قاعدة البيانات المناسبة وكيفية الإفادة منها بطريقة صحيحة .

وعلى الرغم من أن هذه المسوحات قد قامت بفحص المكتبات البحثية الجامعية ، إلا أن النتائج تعتبر فى نظر الباحثين موازية لتلك التى تتم فى أنواع المكتبات الأكاديمية الأخرى ، فالاعتماد على المصادر الإلكترونية وخدماتها والذى بدأ فى عقد التسعينيات من القرن الماضى ، ليس محدوداً بنوع معين من المكتبات مهما كان حجمها وأهدافها .

نتائج الدراسة :

(١) هناك تحول إطارى لمجال استرجاع المعلومات تعكسه المداخل الفيزيائية والمعرفية فالمدخل الفيزيائى Archetypal له جسد واضح من البحث الأمبيريقى فى اختبارات كرانفيلد ، والتى يمكن أن تحتوى على الجوانب الإطارية وإن كانت هذه الجوانب تتميز بالعديد من الغموض Ambiguity لطبيعة الإطار المشمول، كما أن الإنتاج الفكرى للمدخل المعرفى Cognitive تزودنا بصورة واضحة للجوانب التى تبنى عليها الإطار دون أن يكون هناك نموذج أو مثال Exemplar مقبول على نطاق واسع . من أجل ذلك فيذهب الباحثون إلى أن المدخلين ينقصهما الهوية الإطارية الكاملة والتى يمكن أن تخدم كأساس لعلم ناضج متطور Normal Science كما هو الحال فى العلوم الطبيعية كالفيزياء والكيمياء وإنما يكوّن المدخلان إطاراً يدخل ضمن العلوم الاجتماعية كعلم النفس ، أي أن استرجاع المعلومات وهو أحد مجالات علم المعلومات له صفات ما قبل الإطارية Preparadigmatic .

(٢) على الرغم من التزام بحوث استرجاع المعلومات المبكر بالطريقة العلمية Scientific Method فى القياس ، إلا أن هذه البحوث لم تصل إلى إطار كامل لتوجيه البحوث ، وبالتالى فالمشكلات التى كانت موجودة عند تأسيس مجال استرجاع المعلومات ، مازالت تواجه الباحثين المعاصرين عند تحليلهم لمشروعات حديثة مثل تريك TREC = Text Retrieval Conference . واستمرت مشكلة القياسات دون إمكانية السيطرة عليها ، نظراً لطبيعة المادة الموضوعية للمجال وطبيعة الحكم على الصلاحية Relevance وطبيعة المكون المعرفى Cognition and Knowledge وبالتالى فيرى العديد من الدارسين أن مشكلات القياسات فى استرجاع المعلومات شبيهة بمشكلات علم النفس وليس بمشكلات الفيزياء .

(٣) عن مستحدثات محركات البحث Search Engines ومدى استخدامها بذاتها فى الاسترجاع أو مع عوامل البحث الذكية لحل المشكلات التى تعانى من البحث البولينى ، يمكن القول بأن محركات البحث قد أصبحت أكثر الأدوات أهمية فى استرجاع المعلومات على الوب ، ولكن مع تضخم حجم الوثائق وأنواعها على الوب أصبحت بيئة الوب غير محكومة Unconstrained وبالتالى أصبح حساب الاستدعاء Recall عسيراً فى هذه البيئة السريعة التغيير ، كما أن طريقة كرانفيلد التقليدية فى التقييم أصبحت غير ممكنة ، وعن تطور محركات البحث يمكن الإشارة لتجارب تريك الحديثة (TREC-4) والتى تتناول ٣,٢ جيجابايت من النصوص قد أدت إلى أسلوب جديد لتقدير الاستدعاء Recall ، ومع ذلك فإننا نحصل على استدعاء نسبى وليس استدعاء مطلق ، حيث يشمل هذا الاستدعاء المواد الصالحة Relevant Items التى يفشل الباحث فى استرجاعها هذا وتعانى معظم نظم الاسترجاع التقليدية والتى تعتمد على المنطق البولينى من مشكلتين أساسيتين هما تمثيل الأسئلة بمصطلحات غير دقيقة مما يؤدى إلى انخفاض فاعلية الاسترجاع وهذه يمكن حلها عن طريق الشبكات العصبية Neural Networks أما المشكلة الثانية وهى المتصلة بالتكشيف غير المنتظم فقد اقترح لحلها التعلم الاستقرائى Inductive Learning مع الشبكات العصبية أي إتباع المنهجية المهجنة Hybrid Application .

(٤) عن موقع الإنترنت (الوب) فى الخدمة المرجعية الرقمية بالمكتبات الأكاديمية فقد تناولت الدراسة تعريف المكتبة الرقمية وقضاياها البحثية وتطور الخدمة المرجعية الرقمية على الوب ، وقام الباحثون بأربع مسوحات للمكتبات الجامعية خلال الفترة من (١٩٩١-٢٠٠١) وأبرزت الدراسة المسحية أن مصادر المعلومات الرقمية أصبح لها تأثيرها العميق على الوب من كل شئ فى هذه الخدمة ، وأبرزت الدراسة المقارنة بين قواعد البيانات المحملة محلياً وتلك الخاصة بالأقراص المكتنزة والبحث الوسيط وخدمات البحث على الخط المباشر ثم الإتاحة المباشرة للمستفيد على

الوب (الإنترنت) . أى أن الأطر البحثية فى استرجاع المعلومات قد تمت دراستها من النواحى النظرية الفكرية ومن النواحى التطبيقية المتصلة بمحركات البحث وموقع الإنترنت المقارن فى استرجاع المعلومات ، أى أن معظم بحوث استرجاع المعلومات تتعامل داخل الإطارين الفيزيائى والمعرفى ، إلا أن المدخلين يفتقدان إلى الهوية الإطارية التى تشكل أساس علم متطور ناضج Normal Science .

المراجع

١ – أحمد بدر (٢٠٠٢) **الفلسفة والتنظير فى علم المعلومات والمكتبات** – القاهرة : دار غريب للطباعة والنشر والتوزيع .

٢ – أحمد بدر ومحمد فتحى عبد الهادى ونايمان متولى (٢٠٠١) التكشيف والاستخلاص : دراسات فى التحليل الموضوعى ، القاهرة : دار قباء للطباعة والنشر والتوزيع (ص ص ٣٨٤–٣٩٤) .

٣ – نايمان إسماعيل متولى (يناير ٢٠٠٢) الإنترنت وبحوث الوصف والتحليل الموضوعى والاسترجاع فى الإنتاج الفكرى الحديث للمكتبات والمعلومات . **الاتجاهات الحديثة فى المكتبات والمعلومات** مج ١٩ ، ع ١٧ ، ص ص ٩٩–١٢٠ .

4 - Belkin, J. J. (1980) Anomalous states of knowledge as the basis for information retrieval. **Canadian Journal of information Science**. v. 5, 133-143.

5 - Belkin, J. (1990) The cognitive viewpoint in information science. **Journal of Information Science**, v. 16, p 11-15.

6 - Borgman, C. (1999) What are digital libraries ? Competing visions **Information Processing and Management**, v. 35 (3), 277-43.

7 - Brookes, B.C. (1976) A new Paradigm for information science. **The Information Scientist**. v. 10 (3) : 103-111.

8 - Brookes, B. C. (1981) The foundations of information science. Part IV information science : The changing paradigm. **Journal of Information Science**, v. 3, 3-12

9 - Chowdhury, G. G. (March, 1999A) The Internet and information retrieval research : A brief Review. **Journal of Documentation** v. 55 (2), p. 209-225.

10- Chowdhury, G. G. (1999B) **Introduction to modern information retrieval**. London: L.A. Publishing.

11- Chowdhury, G. G. (2002) Digital libraries and reference services:

Present and future. **Journal of Documentation**. v. 58 (3) : 258-283.

12- Cleverdon, C.W. (1971) Design and evaluation of indexing systems. **ARIST**, v. 6: 42-73.

13- Coper, M. D. (2001) Usage patterns of a Web - based catalog. **JASIST**. v. 52 (2) 137-48 .

14- Cortez, E. M. et al (1995) The hybrid application of an inductive learning method and a neural Network for intelligent information retrieval. **information processing and management**, v. 31 (6): 789-813.

15- Croft, W. B. and Thompson, R. H. (1987) IR: A new approach to the design of document retrieval systems. **JASIS**, v. 36 (6), p. 389-404.

16- Curle, D. (1997) Downloading data from the web : you are not in **ASC11** anymore On line, v. 21 (4) 51-58.

17- Db Guide: searching for novices and others. **On line and CD-ROM Review**. v. 21 (4), 242-244.

18- De Mey, M. (1982) **The cognitive Paradigm**. Reidel publishing, Dordrecht.

19- Dodd, D. G. (1996) Grass - roots cataloging and classification: food for thought from www subject oriented hierarchical lists. **Library resources and technical services** 40 (3), 275-286.

20- Doszkocs, T. E., Regiia, J. & Lin, X (1990) Connectionist models and information retrieval. **ARIST**, v. 25, 209-260.

21- Editorial research agenda beyond 2000: **Library & Information Science Research** v. 19 (3), 1997, 209-216.

22- Ellis, D. (1992) The physical and cognitive paradigms in information retrieval research. **Journal of Documentation**, v. 48, 45-64.

23- Elis, D. (1994), Paradigms in information retrieval research **Elis**, v. 54, p. 275-291.

24- Ellis, D. (1996) The Dilemma of measurement in information Retrieval research. **JASIS**, v. 47, 23-36.

25- Ellis, D. (1998) Paradigms and research traditions in information retrieval. **information services & use**, v. 18 (4), p. 225.

fromework for future research. **JASIST**, v. 52 (3), 235-46.

33- Jones, Spark (ed) (1981) **The Cranfield tests in : Information Rerrieval Experiment**. London: Butterworth. p. 256-284.

34- Kowk, K. L. (1989) A neural network for probabilistic information Retrieval **Sigir Forum**, v. 23 (1/2), 21-30.

35- Kowk, K. L., (1990) Application of neural network to information rerrieval. **International Joint Conference on neural nerworks**, II, 623-626.

36- Kuhn, T. S. (1962) **The Structure of scientific revolutions.** Chicago : University of Chicago Press.

37- Lesh, M. The seven ages of information Retrieval . available at: http://commuitybellcore.com/lesk/ages.html.

38- Lowley, S. (March 2000) The Evaluation of WWW search engines. **Journal of Documentation**. v. 56 (2), 190-211.

39- Lynch, C. (1997) Building the infrastructure of resource shraing :

26- Feeney, M. and Karen, M. (1990) **Information technology and the research process**. London: bowker-Saur.

27- Frohman, B. (1990) Rules of indexing: a critique pf mentalism in information retrieval theory. **Journal of Documentation**. v. 46, 81-101.

28- Gatten, J. N. (1991) Paradigm restrictions on interdisciplinary research into librarianship. **College & research libraries**, v. 52: (572-584).

29- Gordon, M. and Pathak, P. (1999) finding information on the Web : the retrieval effectiveness of web search engines. **Information processing & Management**, v. 35 (2) : 141-80.

30- Hofstetter, F. T. (2001) **Multimedia literacy**. Irwin : McGraw Hill khan kushal.

31- Howkins, B. L. and P. Battin (eds) (1998) **The Mirage of Continuity : reconfiguring academic information resourrces for the 21st century**. Washington : Association of American Universities.

32- Jensen, B. & Pooch, V. (2001) A review of web searching studies and a

union catalogs, distributed search, and cross database linkage. **Library trends**, v. 45 (3), 448-461.

40- Masterman, M. (1970) **The nature of a paradigm in** : Criticism and growth of knowledge. Cambridge. Cambridge. Univ. Press.

41- McCarty, M. (1997) Evaluating information on the internet: potential problems for information professionals. **Business information Alert**. v. 9 (2), 11-12.

42- OKAPI and information retrieval research (1997). Special issur of **Journal of Documentation**. v. 53 (1), p. 1-87.

43- Park, T. K. (1993) The Nature of relevance in information retrieval . an empirical study. **Library Quarterly**, 63, 3-48.

44- Popper, K. R. (1972) **Objective knowledge : an Evolutionary approach**. Clarendon Press, Oxford.

45- Schwarty, J. Spring (2002) Internet access and End-User need. **Reference & user services Quarterly**. V. 41 (3), 253-263.

46- Smeation, A. F. and Harman, D. T. (1997) The TREC experiments and their impact on Europe. **Journal of Information Science**, v. 23 (2), 169-74.

47- Tenopir, C. & Lias Ennis (sp 2002) A decade of digital reference **Reference & user services Quarterly.** v 41 (3), 265-273.

48- Younger, J. A. (Winter, 1997) Resource Description in the digital Age. **Library Trends** v. 45 (3), p. 463-481.

49- Vickery, B. C. (1992) knowledge structures in information retrieval. **Journal of Documentation**, v. 48, 326-327.