الترشيح فى بيئة الويب فى ضوء نظريات استرجاع المعلومات : دراسة تحليلية

د. أحمد إبراهيم شاهين
مدرس بقسم المكتبات والمعلومات
كلية الآداب – جامعة بنى سويف

الملخص

يناقش البحث قضية ترشيح المعلومات في بيئة الويب من بعد معلوماتي، وذلك من حيث التعريف على مفهوم ترشيح المعلومات على شبكة الإنترنت وعلاقته بنظم استرجاع المعلومات ونظرياته من حيث أوجه الاتفاق والاختلاف بينهما، وعلاقة الترشيح بالاحتياجات المعلوماتية للمستفيد، مروراً بآليات الترشيح وعلاقتها باسترجاع المعلومات،كما يتعرض البحث لمناقشة قياس معامل الترشيح في ضوء نظريات استرجاع المعلومات، في محاولة لتنظير مفهوم الترشيح في ضوء نظريات الاسترجاع المعروفة في مجال المكتبات والمعلومات وذلك للوصول لمدى مواكبة نظريات الاسترجاع في المجال لمفهوم الاسترجاع وآلياته في ضوء تطبيق ترشيح المعلومات على شبكة الإنترنت.

الكلمات المفتاحية: نظم استرجاع المعلومات المفتوحة، نظريات استرجاع المعلومات، ترشيح المعلومات

١/١تمهيد:

أثرت ظهور الإنترنت على التعاطي مع الحاجات المعلوماتية للمستفيدين، حيث اتسمت بالزخم والفوضى المعلوماتية التي حاولت نظم الاسترجاع المفتوحة كمحركات البحث وادلة البحث ومن بعدها البوبات السيطرة عليها بهدف الاستثمار المعرفي للزخم المعلوماتي المتجدد لتحقيق أعلى معدلات الافادة للمستفيد، وظل الأمر

يشغل بال المعلوماتيين والمتخصصين في علوم الحاسب ولكن في معزل عن بعضهم البعض، ولكنها التقت في نهاية الأمر في تقنيات وبرمجيات تعيد صياغة مفهوم معلوماتي، ألا وهو الترشيح الذي طبقته المكتبات في خدمات البث الانتقائي للمعلومات، وقد ابرزت آليات وبرمجيات الترشيح العلاقة الوثيقة بين الترشيح واسترجاع المعلومات على مستوى النظرية والتطبيق. لذا تسعى هذه الدراسة للتأصيل العلمي للعلاقة بين مفهوم الترشيح واسترجاع المعلومات على المستوى النظري والتطبيقي.

٢/١ أهمية البحث

تنبع اهمية البحث من ما يلي:

١. ندرة الدراسات العربية في مجال ترشيح المعلومات وعلاقته بنظم الاسترجاع.
٢. حدوث تطور في طبيعة المعلومات ومصادرها ونظم الإتاحة الرقمية سمح بوجود اساليب جديدة في استرجاع المعلومات.
٣. ظهور متغيرات جديدة في بيئة نظم الاسترجاع الرقمية على الويب تستلزم إعـــادة النظر في نظريات استرجاع المعلومات المعروفة في المجال.

٣/١ مشكلة البحث

تنبع مشكلة البحث الحالي من بروز بيئة الويب في الآونة الأخيرة كنظام استرجاع مفتوح وانتشارها بمتغيرات جديدة لم تكن متاحة وقت وضع نظريات استرجاع المعلومات المعروفة في مجال المكتبات والمعلومات مما أظهر عدم مواكبة هذه النظريات لقياس معاملات الاسترجاع في هذه البيئة. لذا يسعى هذا البحث للتوصل إلى آلية لقياس معامل ترشيح المعلومات في بيئة الويب في ضوء نظريات استرجاع المعلومات لتحديد مدى تأثيره على نتائج الاسترجاع.

١/ ٤ الفرضية

ينطلق هذا البحث من فرضيتين أساسيتين هما:

- أن ترشيح المعلومات يمثل شكلا جديدا من أشكال أسترجاع المعلومات تطبق آلياته في نظم استرجاع المعلومات المفتوحة.
- أنه يمكن قياس معامل ترشيح المعلومات بنظم الاسترجاع المفتوحة في بيئة الويب استناداً على نظريات استرجاع المعلومات المعروفة في المجال.

٥/١ الأهداف

١. إلقاء الضوء على علاقة مفهوم الترشيح بنظريات استرجاع المعلومات.
٢. محاولة لتنظير مفهوم الترشيح في ضوء نظريات الاسترجاع المعروفة في مجال المكتبات والمعلومات.
٣. وضع تصور لقياس معاملات الاسترجاع في البيئة الرقمية في ضوء معطيات هذه البيئة.
٤. رسم تصور مقترح لما تمثله النظرية الجديدة لاسترجاع المعلومات المتضمنة لمعيار الترشيح بين معايير الحكم على نظم استرجاع المعلومات المفتوحة.

٦/١ حدود الدراسة

الحدود الموضوعية: تهتم الدراسة بتحليل نظريات استرجاع المعلومات في ضوء آليات ترشيح المعلومات من زاوية معلوماتية.

٧/١ منهج الدراسة وأدواتها

تعتمد هذه الدراسة على المنهج الوثائقي في رصد المؤشرات التي وردت في نظريات استرجاع المعلومات المعروفة في مجال علم المكتبات والمعلومات. وسيتم التعرض بالتحليل هنا لنظريات استرجاع المعلومات وتطبقاتها في بيئة الويب باعتبارها نظام استرجاع مفتوح في ضوء متغير ترشيح المعلومات، وذلك من خلال استعراض أدبيات الموضوع للتعرف على العلاقة بين ترشيح المعلومات ونظم استرجاع المعلومات المفتوحة.

اعتمد الباحث على أداة البحث الوثائقي في القراءات النظرية للعديد من الدراسات التي تناولت هذا الموضوع.

٨/١ أدبيات الإنتاج الفكري

استعراض الباحث أدبيات الموضوع في أدلة الإنتاج الفكري في مجال المكتبات والمعلومات وقواعد البيانات العالمية والدوريات المتخصصة التي تحتوي على النصوص الكاملة، وكذلك محركات البحث بمصطلحات الدراسة، وهي:

- استرجاع المعلومات Information Retrival
- ترشيح المعلومات Information Filtering

لاحظ الباحث أن أدبيات الإنتاج الفكري العربي والأجنبي لم تتناول العلاقة بين الترشيح واسترجاع المعلومات من زاوية معلوماتية، غير أن الإنتاج الفكري الأجنبي عالج العلاقة بين المفهومين من زاوية تقنية. وكان أهم ما افرزه الإنتاج الفكري مرتبط بموضوع الدراسة:

- **فاتن سعيد بامفلح. إضفاء الطابع الشخصي على عمليات البحث واسترجاع المعلومات.- مجلة مكتبة الملك فهد الوطنية مج ١٦ ، ع٢، رجب - ذو الحجة ١٤٣١ هـ / يونيو - نوفمبر ٢٠١٠ م.**

وتهدف الدراسة إلى التعرف إلى مفهوم البحث الشخصي من الناحية النظرية، وتحديد مدى اختلاف تطبيق تلك المقاربة بين نظم المعلومات المختلفة، وإيضاح مدى حماية الخصوصية في ظل تطبيق البحث الشخصي. وقد استخدمت الباحثة المنهج الوصفي التحليلي على عينة قصدية تضم مواقع بعض نظم المعلومات المتاحة عبر شبكة الإنترنت، للتعرف إلى تطبيقات أسلوب البحث الشخصي في كل موقع قاعدة بيانات ERIC، ومحرك البحث Google، وموقع Amazon، وكانت أهم نتائج الدراسة ان المواقع تقدم خدمات ذات صبغة شخصية.

- **Palme, Jacob. Information filtering. - July 1998. - Accessed December 2004. - Available At: http://www.dsv.su.se/select/information-filtering.html.**

نشرت هذه المقالة لأول مرة بتاريخ ١ يونيو ١٩٩٨ في مؤتمر ITS'98 conference، ثم حُدّثت بتاريخ ٣ يوليو ١٩٩٨ من قبل مؤلفها Jacob Palme الذي يعمل بقسمِ علوم الحاسب والنظم، جامعة Stockholm. ويدور البحث حول أفضل الطرقِ لإيجاد المعلومات القيمة على الإنترنت، وتجنب المواد غير القيمة؛ مما يحسن من معدلات الإفادة من شبكة الإنترنت، ويقدم المقال نظرة عامة حول طرائق الترشيح الرئيسية، وقضاياه، وكذلك الترشيح الاجتماعي، وأثر الترشيح في نتائج البحث، وأخيراً يتعرض للآراء المؤيدة والمعارضة للترشيح.

- **Belkin, Nicholas J, Bruce Croft. Information filtering and information retrieval: Two sides of the same coin?.- Communications of the ACM Dec 1992 v35 n12 p29(10)**

صممت نظم ترشيح المعلومات للتعامل مع البيانات غير المهيكلة وشبة المهيكلة، عكس نظم إدارة قواعد البيانات التي صممت للتعامل مع البيانات المهيكلة، كما صممت نظم الترشيح للتعامل مع امعلومات النصية، ولكنها قد تتمكن من التعامل مع ملفات الصور، والصوت، والفيديو، وباقي انواع البيانات المتاحة في نظم المعلومات. كما تتعامل نظم الترشيح مع الكم الضخم من البيانات المضافة والمرسلة عن بعد، ويتم الترشيح بناء على توصيف الاحتياجات المعلومات الشخصية أو لمجموعة من الاشخاص، أو بناء على سمات المستفيد والذي يمثل احتياجاته المعلوماتية طويلة الآجل. ويهدف الترشيح لاستبعاد المعلومات التي لاتمثل اهتمام المستفيد من البيانات المسترجعة، واسترجاع المعلومات التي تمثل اهتماماته فقط، لذا سيعرض البحث لنماذج ترشيح واسترجاع المعلومات، والدراسات في مجال استرجاع وترشيح المعلومات.

- **URI HANANI, BRACHA SHAPIRA and PERETZ SHOVAL. Information Filtering: Overview of Issues, Research and Systems .- User Modeling and User-Adapted Interaction.- Volume 11, Number 3 (august 2001) .- 203-259.**

يناقش البحث التدفق المعلوماتي المنشور عبر وسائط المعلومات الإلكترونية وإشكالية المستفيد في التعامل معه، وضرورة الحاجة لأدوات لإدارة موارد المعلومات، ويعد ترشيح المعلومات أحد منهجيات التعامل مع التدفق المعلوماتي الضخم. حيث يهدف الترشيح لتوفير المعلومات فقط التي يرغبها المستفيد. كما ناقش تطور نظم وبرمجيات الترشح لتلائم الاحتياجات البحثية عبر الإنترنت، مثل البريد الإلكتروني، والمجموعات الاخبارية، وخدمات التجارة الإلكترونية بهدف تقديم خدمات ترشيح شخصية وجماعية من خلال استبعد المواد التي تهم المستفيد، واستبعاد المواد غير الملائمة للأطفال. كما تعرض البحث لمفاهيم وآليات وتقنيات الترشيح، وتطبيقات مفاهيم استرجاع المعلومات، والذكاء الاصطناعي، والدراسات السلوكية في ترشيح المعلومات، هذا إلى جانب نظم الترشيح المختلفة المبنية على نموذج المستفيد. كما اوضح الباحثان أوجه الاتفاق والاختلاف بين الترشيح واسترجاع المعلومات.

٩/١ المصطلحات الإجرائية

لأغراض هذه الدراسة سيتم وضع التعريفات للمصطلحات التالية:

- **نظم الاسترجاع المفتوحة**: هي نظم الاسترجاع المبنية على قواعد بيانات والتي تقوم بتجميع المعلومات بشكل آلي وحفظها وتحديثها واسترجاعها، حيث تتلقى قاعدة البيانات مدخلات متمثلة في مصادر المعلومات المتاحة في بيئة الويب، ليتم معالجتها وتقدم مخرجات بناء على الاستفسارات البحثية المقدمة من المستفيدين، وتتم عملية الجمع والحفظ بصفة مستمرة بواسطة برمجيات تتولى الحصر والإضافة لقاعدة بيانات النظام نظراً للنمو المتزايد لمصادر المعلومات في بيئة الويب؛ كما تقوم هذ البرمجيات بحذف مصادر المعلومات التي انتهت استضافتها في بيئة الويب من قاعدة بيانات النظام واستبعادها.
- **ترشيح المعلومات**: هي عملية استبعاد لا نهائي للمعلومات التي لا تمثل احتياجات المستفيد الحالية، والتي يمكن استرجاعها لاحقاً في حال تغير احتياجات المستفيد البحثية ووقعها ضمن دائرة اهتماماته البحثية.

- **معامل الترشيح**: هي نسبة التأثير على نتائج الاسترجاع قبل تطبيق معايير الترشيح التي يحددها المستفيد وبعدها.

١/٢ مصادر المعلومات المتاحة على الإنترنت والترشيح

أصبحت شبكة الإنترنت في السنوات الأخيرة مصدرا هاما للإنتاج الفكري العالمي، وعامل مؤثر في إعادة صياغة أنماط توزيع المعلومات والإفادة منها. والمصادر المتاحة على هذه الشبكة تفوق في عددها تلك الموجودة في مكتباتنا الكبرى. ومازال هناك الكثير غير المعروف عن طبيعة المعلومات المتاحة على شبكة الويب ومجوعات الوثائق التي تشتمل عليها. ولكن هناك شيء واحد هو الواضح في هذا الصدد هو أن الويب عبارة عن مجموعة من الوثائق التي أسهم بها مؤلفين أو ناشرين يملكون برمجيات النشر الرقمي التي تمكنهم من النشر على الويب، ومن المعروف أن الويب ليس بها سياسة اختيار كتلك السياسات الدقيقة الموضوعة للاختيار بالمكتبات؛ حيث أن القرارات الخاصة بإضافة أي وثائق على الشبكة وتحديثها يسودها الارتباك؛ فهذا الوثائق المنشورة على الويب تشتمل على الإيجابي منها والسلبي.[١]

ومع النمو المتزايد للوثائق على شبكة الإنترنت وشبكات المعلومات الأخرى، فإن عدم استخدام البحث بالترشيح الآلي سيؤدي إلى تضخم شبكات المعلومات في السنوات القريبة. وذلك بسبب انخفاض تكلفة الاتصال بالشبكة، والسهولة في الوصول إلى معلومات جديدة. وقد جعل النمو الحالي المتزايد للشبكة الويب العالمية استخدام الترشيح الآلي من الموضوعات المثيرة لاهتمام الباحثين، وذلك بالرغم من مشكلة تجميع معلومات في هذا الموضوع مما صعب من مهمة إجراء بحث رئيسي عن تقنيات ترشيح المعلومات.[٢]

---

[1] أونيل، إدواد ت . خصائص المعلومات المتاحة عبر شبكة العنكبوت العالمية / ترجمة عبد الرحمن فراج. – دراسات عربية في المكتبات وعلم المعلومات. – مج ٥، ع ٣ (سبتمبر ٢٠٠٠). ص ١٩٧.

[2] Oard, Doug . "Information Filtering Defined" . - Last modified: Tue Dec 12 15:33:26 1995 . – Access at April 2011 . – Available at: http://www.enee.umd.edu/medlab/filter/filter_definition.html

فقد أصبح ضروريا أن يكون هناك طرق آلية لترشيح المعلومات مع الإتاحة المتزايدة للمعلومات في الشكل الإلكتروني. فقد أنتجت المنظمات البحثية كميات كبيرة من المعلومات، تتضمن مذكرات تقنية وإدارية، وإعلانات اجتماعات ومؤتمرات، ومقتطفات من الاجتماعات.

وسيكون من الصعب مع هذا الكم الكبير من المعلومات إبقاء المستفيدين على علم بكل الأعمال ذات العلاقة باهتماماتهم. وعلاوة على ذلك، سيكون جزء صغير فقط من المعلومات المتاحة ذو علاقة فعلية باهتمامات مستفيد معين يعمل ضمن منظمة تغطي قطاعات مختلفة. وهكذا، فأن هناك مشكلة في تحديد ما هي المعلومات التي تمثل مجالات اهتمام للمستفيد، وذلك لتقليل مقدار البحث في المعلومات غير ذات العلاقة باهتماماته.[٣]

لذا يرى الباحث أن ترشيح المعلومات إلى جانب حتميته له أهميتان؛ أولهما تتعلق بدور الترشيح في الإفادة من مصادر المعلومات الضخمة والمتنوعة المتاحة على شبكة الإنترنت، وذلك من خلال ترشيح المعلومات وفقاً لاحتياجات المستفيد فيما يشبه خدمة البث الانتقائي للمعلومات، وتتعلق ثانيهما بحماية المستفيد من أية أضرار قد يتعرض لها أثناء تصفحه للإنترنت وذلك باستخدام التقنيات المتنوعة للقيام بذلك بغرض تحقيق هدف واحد وهو تلبية احتياجات المستفيد من البيئة الرقمية لشبكة الإنترنت.

٢/٢ العلاقة بين الترشيح واسترجاع المعلومات من حيث المفهوم

يعد مفهوم استرجاع المعلومات من المفاهيم المستقرة في علم المكتبات والمعلومات، بينما مفهوم ترشيح المعلومات استقر في علوم الحاسبات لخضوعه للدراسة والتناول كنظام وكبرمجيات منذ منتطف التسعينيات من القرن الماضي؛ إلا أنه لم يحظى بنفس الاهتمام والتناول المعلوماتي من المتخصصين في علم المكتبات

[3] Foltz, Peter W., Dumais, Susan T. Personalized Information Delivery: An Analysis of Information Filtering Methods . - Access at November 2011 . – Available at: http://www-psych.nmsu.edu/~pfoltz/cacm/cacm_fn.html

والمعلومات إلا في بدايات القرن الحالي، لذا سيعرض الباحث لمفهومي الترشيح واسترجاع المعلومات في محاولة للتأصيل المعلوتي للعلاقة بين المفهومين.

فقد كانت هناك العديد من المفاهيم التي وقعت ولفترة طويلة تحت مصطلح استرجاع المعلومات منها أنظمة المعلومات والحاسبات وخزن واسترجاع المواد المكتبية. وفي مجال المكتبة الرقمية تستخدم العديد من تقنيات استرجاع المعلومات، من ضمنها بحث الميتاداتا metadata، والبحث عن النصوص الكاملة للوثائق، وبحث المحتوى عن أنواع البيانات الأخرى.[٤] في حين أن المصطلحات المتصلة بترشيح المعلومات متغيرة وأحياناً غير واضحة نظراً للطبيعة المتغيرة لهذا المجال.[٥]

يعرف نظام استرجاع المعلومات بأنه "يقود المستفيد نحو الوثائق التي تلبي احتياجاته المعلوماتية"، وتصف العديد من التعريفات نموذج استرجاع المعلومات الذي يبدأ باستفسار بحثي من المستفيد لنظام المعلومات كقاعدة بيانات بهدف الوصول لمصادر المعلومات ذات الصلة بمجال اهتمامه من بين العديد من مصادر المعلومات المتاحة في قاعدة البيانات، وذلك من خلال المصطلحات الكشفية لمصادر المعلومات المخزنة بقاعدة البيانات.[٦]

ويرتبط مفهوم ترشيح المعلومات Information Filtering بمفهوم استرجاع المعلومات Information Retrieval، باعتبار أن ترشيح المعلومات جزء من عملية التحكم في المعلومات المستدعاة أو المسترجعة.[٧] ويعتبر المفهوم الاصطلاحي

---

[4] Nabil R. Adam …[others] . "Digital Library Technical Committee" . – Access at April 2011 . – Available at: http://cimic3.rutgers.edu/ieee_dltf.html

[5] URI HANANI, BRACHA SHAPIRA and PERETZ SHOVAL. Information Filtering: Overview of Issues, Research and Systems .- User Modeling and User-Adapted Interaction.- Volume 11, Number 3 (august 2001) .- 205.

[6] Kuflik , Tsvi and Shoval, Peretz. Generation of User Profiles for Information F i l t e r i n g -Research Agenda.- Access at April 2011 . – Available at: http://www.hmhsieh.idv.tw/ec/3b_1_06.doc

[7] سيد ربيع سيد. محركات بحث الصور الثابتة على الإنترنت. – الرياض: مكتبة الملك فهد الوطنية،٢٠٠٧.- ص ٢٢٨.

لاسترجاع المعلومات Information Retrieval (IR) وثيق الصلة جدا بترشيح المعلومات Information Filtering (IF) في أن كلاهما يهدفان لاسترجاع المعلومات ذات العلاقة بما يريد المستفيد، بينما يقلل كمية المعلومات غير ذات العلاقة المسترجعة. [٨]

وقد ظهرت علاقة مفهوم الترشيح مع نظم الاسترجاع في العديد من التعريفات، ومن بينها تعريف Nicholas Belkin الذي يرى أن ترشيح المعلومات يساعد المستفيدين على مراقبة وتنقية المعلومات، أي انه بشكل أكثر تحديدا يتعلق بالتحكم في طبيعة المعلومات وموضوعاتها الواردة إلى المستفيدين"، وفي حين يرى سيد ربيع أن ترشيح المعلومات في المعنى العام له يعني "البرنامج أو الإجراءات أو السياسات التي يضعها مصمم مواقع الويب أو مستخدم الحاسب الآلي للتعامل مع محتويات محددة سلفا من المعلومات المسترجعة أمام المستفيد". [٩]

في حين أن Doug Oard يرى الهدف من نظم ترشيح المعلومات هو الفحص والتدقيق بفاعلية في الكم الضخم والمتزايد من المعلومات لتقديمها إلى المستفيد أو المستفيدين الراغبين في إشباع رضائهم المعلوماتي أو متطلباتهم المعلوماتية. [١٠]

ويرى الباحث أن المفاهيم والتعريفات السابقة ابرزت محاور الاتفاق والاختلاف بين مفهومي الترشيح واسترجاع المعلومات، تتركز محاور الاتفاق في الهدف المتعلق بتلبية احتياجات المستفيد من خلال التحقيق، والبيئة المتعلقة بالتعامل مع نظام استرجاع يضم بيئة معلوماتية ضخمة، والآلية المتمثلة في استفسار المستفيد المعبر عن احتياجاته المعلوماتية، وتتضح نقاط الاختلاف بين المفهومين في تركيز الترشيح على الاحتياجات الشخصية والسمات المعلوماتية للمستفيد.

لذا وفي ضوء ما سبق، يرى الباحث أن ترشيح المعلومات هي عملية استبعاد لا نهائي للمعلومات التي لا تمثل احتياجات المستفيد الحالية، والتي يمكن استرجاعها لاحقاً في حال تغير احتياجات المستفيد البحثية ووقعها ضمن دائرة اهتماماته البحثية.

---

[8] Foltz, Peter W., Dumais, Susan T. Loc. Cit.

[9] سيد ربيع سيد. مصدرسابق .- ص ٢٢٨.

[10] Oard, Doug . Loc. Cit.

١/٢/٢ الترشيح واسترجاع المعلومات في ضـــوء الاحتياجـــات المعلوماتيـــة للمستفيد

يرتبط ظهور مفهوم الترشيح بالاحتياجات المعلوماتية للمستفيد، فقد كانت الحاجة ملحة لظهور الترشيح في ظل التطور النوعي والكمي في المجتمع المعلوماتي الذي يتعامل معه المستفيد على مستوى شبكة الإنترنت والمكتبات الرقمية، لذا استوجب الأمر تطبيق مفاهيم جديدة للتعاطي مع المتغيرات المعلوماتية بهدف تلبية الاحتياجات المعلوماتية للمستفيد في ظل الزخم المعلوماتي المتاح دون ضوابط معيارية كتلك المتوفرة في سوق النشر التقليدي، لذا سيلقى الباحث الضوء على علاقة الترشيح بنظم استرجاع المعلومات في بيئة الويب من حيث وحدة الهدف ألا وهو تلبية الاحتياجات المعلوماتية للمستفيد.

أصبحت شبكة الإنترنت في السنوات الأخيرة مصدرا هاما للإنتاج الفكري العالمي، وعامل مؤثر في إعادة صياغة أنماط توزيع المعلومات والإفادة منها. والمصادر المتاحة على هذه الشبكة تفوق في عددها تلك الموجودة في مكتباتنا الكبرى. وتضم الويب مجموعة من الوثائق التي أسهم بها مؤلفين أو ناشرين يملكون برمجيات النشر الرقمي التي تمكنهم من النشر على الويب، ومن المعروف أن الويب ليس بها سياسة اختيار كتلك السياسات الدقيقة الموضوعة للاختيار بالمكتبات؛ حيث أن القرارات الخاصة بإضافة أي وثائق على الشبكة وتحديثها يسودها الارتباك؛ فهذا الوثائق المنشورة على الويب تشتمل على الإيجابي منها والسلبي.[11]

ويتمثل تفاعل المستفيد مع نظم الاسترجاع في كيفية تحديده لاستفساره؛ وكيفية ترجمته للإجابات التي يقدمها له النظام، أما عملية صياغة الاستفسار فليست بالأمر البسيط، إلا في الحالات التي تكون فيها الاستفسارات بسيطة؛ وحتى في حالة تحديد الاستفسار بطريقة جيدة، فإن المستفيد سيجد أن النظام يقدم له كما كبيرا من صفحات

[11] أونيل، إدوارد ت . خصائص المعلومات المتاحة عبر شبكة العنكبوت العالمية / ترجمة عبد الرحمن فراج. – دراسات عربية في المكتبات وعلم المعلومات. – مج ٥، ع ٣ (سبتمبر ٢٠٠٠). ص ١٩٧.

الويب التي قد تصل إلى الآلاف، وعليه أن يتعامل معها بطريقة فعالة. ويذكر أن لأحد المتخصصين عبارة دالة على الواقع الذي نعيشه، حيث وصف الوضع الحالي بقوله" :

نحن نغرق في كتلة من البيانات، في حين أننا نموت من العطش للمعلومات

"we are drawing in the mass of data, while we dying of thirst for Information"[١٢]

ويمكن أن يقاس نجاح استرجاع المعلومات بالنسبة المئوية للمعلومات المسترجعة ذات العلاقة وغير ذات العلاقة. ومن الصعب التحديد بدقة وبشكل كمي فعالية استرجاع المعلومات؛ ولكن يمكن فقط للمستعمل الفرد أن يقرر ما هو المفيد. وتتضمن تقنيات تحسين مستوى وفاعلية وتأثير عملية الاسترجاع على الإعداد للوثائق للحصول على الميتاداتا الإضافية قبل تخزينها في قاعدة بيانات الوثائق.[١٣]

لذا فإن هناك حاجة كبيرة إلى توفير الوسائل التي تتيح الوصول إلى الاحتياجات المستفيدين مع التركيز على جودة المعلومات، أي الكيف وليس الكم، حتى يتمكن المستفيد من الحصول على المعلومات ذات الصلة الوثيقة باستفساراته .ولعل ذلك ما دفع القائمين على نظم استرجاع المعلومات إلى السعي بصورة مستمرة نحو تطوير أساليب البحث والاسترجاع ليتمكن المستفيد من الحصول على احتياجاته من المعلومات بدقة أكبر في ظل ذلك الفيض الهائل من المعلومات، وتنوع وسائطها، ولغاتها، ومواقعها، وصعوبة صياغة الاستفسارات التي تصل بالمستفيد إلى ما يحتاج إليه في ظل تعقد احتياجاته من جهة، وتعقد الموضوعات وتداخلها من جهة أخرى. وهناك كثير من

[12] فاتن سعيد بامفلح. إضفاء الطابع الشخصي على عمليات البحث واسترجاع المعلومات.- مجلة مكتبة الملك فهد الوطنية مج ١٦ ، ع٢، رجب - ذو الحجة ١٤٣١ ه / يونيو - نوفمبر ٢٠١٠ م. ص ٦١

[13] Nabil R. Adam …[others] . Loc. Cit.

أساليب وتقنيات واتجاهات البحث التي تم تطويرها في الآونة الأخيرة، ومن بينها :البحث الدلالي، والبحث الاجتماعي، والبحث الشخصي، والتنقيب عن المعلومات.[١٤]

وقد أكد Broder ان الويب قد حملت معها انماطا جديدة لحاجات المستفيدين في التعامل مع المعلومات سواء كان على صعيد البحث أوالاسترجاع، وقد امتدت هذه الانماط لتشكل فئات جديدة من الاستفسارات المعلوماتية لدى المستفيدين، ففي سياق االويب، لا تقتصر الحاجة من وراء الاستفسار على ان تكون ذو طبيعة معلوماتية فحسب بل تمتد لتشمل ثلاث اقسام اخرى من حاجات المستفيد، وهي:

- **الحاجة الملاحية**: يعد نمطا حديثا من الاستفسارات فالغرض منه هو الوصول إلى موقع أو مصدر معلومات معين لدى المستفيد وأن الحاجة من وراء هذا الاستفسار تتمثل في أن المستفيد ربما قد زار أو تعرض لهذا الموقع من قبل او على فرضية ان مثل هذا الموقع يمكن ان يكون موجودا على الويب. إن هذا النمط من الحاجات او البحث كان يشار اليه سالفا بالبحث عن "المفردة المعروفة" بمعنى ان المستفيد يكون على دراية وعلم بأن هذه المفردة (سواء كانت هذه المفردة وثيقة او ملف صوتي او موقع ..الخ) موجودة على الويب.
- **الحاجة المعلوماتية**: يعد هذا النمط تقليديا إلى حد ما فالغرض منه لايختلف كثيرا عن الغرض من الحاجات المعلوماتية من نظم الاسترجاع التقليدية وهو ايجاد المعلومات التي يفترض ان تكون متاحة على الويب في صورة ثابتة ولكن سمة امتياز امتازت به الويب تمثل في قدرتها على المزج بين البيانات المتاحة من خلال المصادر المختلفة في صورة اقرب لتحقيق التكاملية بين مصادر الويب فمثلا من خلال البحث عن مدينة الرياض في بعض محركات البحث تكفل خوارزميات هذه المحركات القدرة على توفير مصادر المعلومات التي تتناول الرياض ثم تستطرد لتسترجع اسماء فنادق الرياض والمعالم الاثرية بها... الخ .
- **الحاجة الاجرائية**: والغرض من طبيعة هذه الحاجة هو الوصول إلى مواقع تتوافر فيها سمة التفاعل سواء كان بين المستفيدين بعضهم البعض (كالشبكات

[14] فاتن سعيد بامفلح. مصدر سابق. ص ٦١

الاجتماعية) او تفاعل المستفيدون مع الالة (كتحميل الملفات والتنزيل الهابط وكالمعاملات التجارية مع البنوك).[15]

ويرى الباحث أن الاشكاليات السابقة المتمثلة في التضخم المعلوماتي في نظم استرجاع المعلومات المفتوحة التي تمثلها بيئة الويب من جانب، وتعقد حاجات المستفيد المعلوماتية من جانب أخر،ويواكب هذه الاشكاليات توافر تقنيات وآليات مساعدة لنظم استرجاع المعلومات متمثلة في آليات ترشيح المعلومات تهدف لتقديم خدمة معلوماتية ذات طابع شخصي للمستفيد.

ويؤيد رأي الباحث ما توصلت إليه **فاتن بامفلح**، والتي ترى أن ترشيح المعلومات يعد واحد من الطرق المختلفة التي تجعل المعلومات المقدمة للمستفيد دائما ما تكون اقرب إلى تلبية احتياجات المستفيد المعلوماتية، هذا إلى جانب خدمات تحمل الصبغة الشخصية والتي من شانها خلق نوع من ولاء المستفيد للموقع فضلا عن تحقيقها لدرجة اعلى من الفاعلية في نتائج البحث ومنها حجب مواقع معينة.[١٦] وتعد نظم ترشيح المعلومات مؤشراً على الاحتياجات المعلوماتية طويلة المدى للمستفيدين والمتمثلة في سمات المستفيد User profiles، وتعتمد جودة قياس مؤشرات سمات المستفيد على أداء نظم ترشيح المعلومات.[١٧] حيث تساعد نظم ترشيح المعلومات المستفيدين على استبعاد المعلومات غير ذات العلاقة باحتياجاتهم المعلوماتية وذلك باسترجاع المعلومات ذات العلاقة باحتياجاتهم المعلوماتية، ويعمل الترشيح كوسيط بين مصدر المعلومات والمستخدم النهائي.[١٨]

---

[15] مؤمن سيد النشرتي. التحديات التي تواجه خوارزميات محركات البحث في استرجاع المحتوى العربي على الشبكة العنكبوتية العالمية: دراسة مسحية تحليلية.- ورقة بحثية مقدمة لمؤتمر المحتوى العربي على الانترنت.- ٣-٥ اكتوبر ٢٠١١ جامعة الامام محمد بن سعود الاسلامية.

[16] فاتن سعيد بامفلح. مصدر سابق. ص ٦٢

[17] Kuflik , Tsvi and Shoval, Peretz. Loc. Cit.

[18] Zhuhadar Leyla, Nasraoui Olfa, Robert Wyatt, "Automated Discovery, Categorization and Retrieval of Personalized Semantically Enriched E-learning Resources ", The Third IEEE International Conference on Semantic Computing (ICSC-09), Berkeley, CA, USA - September 14-16, 2009.- P.P 1-2.

ولا يعد مفهوم ترشيح المعلومات من المفاهيم الجديدة على المجتمع المعلوماتي، ولا يعد أحد محددات الاختيار للوثائق الإلكترونية. فعندما يتم قراءة نصوص ورقية يتبادر لذهن المستفيد مفهوم ترشيح المعلومات. فالمستفيد يشتري دوريات محددة فقط لأن الدوريات الباقية قد تحتوي على معلومات غير هامة وغير ذات علاقة باهتماماته، وبهذه الطريقة يقوم المستفيد بترشيح البعض من الكمية الكبيرة من المعلومات التي يمكننا الوصول إليها. وحتى ضمن الدورية الواحدة يختار المستفيد المقالات التي يرى أنها ذات علاقة باهتماماته، وهكذا فأن ترشيح المعلومات هو أمر يقوم به المستفيد بشكل مستمر لأي معلومات يرغب في اقتنائها. ومع الإتاحة الإلكترونية للمعلومات، فأنه لم تعد هناك حاجة ليقوم المستفيد بعمل الترشيح بنفسه ولكن من الممكن أن يتم الترشيح آليا من خلال النظام الذي يقدم المعلومات.[١٩]

وقد تم تطبيق ترشيح المعلومات في العديد من المجالات مستخدما رؤى تقنية متنوعة. وتمثلت الطرق الأولية في قطاع المكتبات والمعلومات في خدمات الإحاطة الجارية بالمكتبات البحثية والمتخصصة لإعلام المستفيدين بالمعلومات الجديدة التي تقع في مجال اهتمامهم. ومع الوقت أصبحت تدعى هذه الخدمة بالبث الانتقائي للمعلومات Selective Dissemination of Information (SDI، وقد تبنت بضعة أنظمة حديثة هذا الاسمِ الوصفيِ لعملية الترشيح، وبوجه عام فإن سبب الاهتمام بترشيح المعلومات نتج من الإقبال البحثي الحالي مع ظهور المكتبات الرقمية على الأقل في جزء من العمليات التقليدية. [٢٠]

ويجب التمييز بين تجميع المعلومات information collection وترشيح المعلومات information filtering . حيث نجد في بعض المجالات - مثل أخبار Usenet - أن جهد التجميع للمعلومات أقل ما يمكن، وذلك لأن المعلومات تأتي للمستفيد. وفي مجالات أخرى - مثل الشبكة العالمية – نجد أن جهد التجميع يمكن أن يكون كبيراً لأنه لا توجد آلية لتجذب المعلومات الجديدة انتباه نظام الترشيح. ويرى Doug Oard أن مجال تجميع المعلومات من المجالات المثيرة للاهتمام ولكن

[19] Foltz, Peter W., Loc. Cit.
[20] Oard, Doug . Loc. Cit

الغرض الأساسي من تجميع المعلومات هو الإعلام بها، وعند الإعلام بها تبدأ مشاكل ترشيح المعلومات، لأن مشاكل الترشيح تبدأ عندما يتمكن المستفيد من الوصول إلى معلومات جديدة. [٢١]

وقد ظهر ترشيح المعلومات في الآونة الأخيرة كمنهجية لتقديم المعلومات ذات العلاقة باحتياجات المستفيدين. وتغطي نظم ترشيح المعلومات قطاعات موضوعية عريضة (مجالات معرفية عريضة)، وتكنولوجيات، ومنهجيات يستخدمها المستفيدون أثناء إجراء الاستفسارات البحثية للوصول للمعلومات التي تمثل إحتياجاتهم الفعلية. وتنطبق نظم ترشيح المعلومات على البيانات غير المنظمة أو شبه المنظمة مثل الوثائق والبريد الإلكتروني، لذا فأن نظم الترشيح تعالج كمية كبيرة من البيانات، وتتعامل مع البيانات النصية بصفة أساسية، وتعتمد على سمات المستفيد، وتهدف إلى استبعاد البيانات غير ذات الصلة باهتمامات المستفيد من نتائج الاسترجاع. [٢٢]

وقد كان هناك اهتمام بحثي كبير في المؤتمر السنوي لاسترجاع النص Text REtrieval Conference (TREC) الذي عقد في جامعة Maryland تحت رعاية قسم هندسة الحاسبات والالكترونيات Department of Electrical and Computer Engineering عام ١٩٩٥ " بأسلوب التقييم التحكمي Controlled Evaluation Methodology" المبني على استعمال مجموعة نصوص قياسية. وقد كانت أهم نتائج المؤتمر أن ترشيح المعلومات يعد أحد مهام الموجه "Routing"، مما تسبب في مزيد من الخلط والتشويش في مفهوم المصطلح في حقل المعلومات. وقد أهتم المؤتمر مؤخرا بالبحوث الخاصة بمسار الترشيح والتي تتبنى أسلوب تقييم evaluation methodology ولكنها تتوافق مع مفهوم وتعريف الترشيح الذي تبناه Doug Oard مسبقا. مما سيؤدي في النهاية إلى توافر أنظمة تجارية لترشح المقالات والأخبار

[21] Op. Cit.
[22] URI HANANI, BRACHA SHAPIRA and PERETZ SHOVAL. Loc. Cit. .- 203-204.

ومصادر المعلومات المتخصصة الأخرى، ومن المحتمل أن تطبق تقنيات الترشيح في المستقبل على مجالات أخرى مثل الصور والصوت والفيديو. [٢٣]

ولا تقتصر السمات المذكورة سابقا على نظم ترشيح المعلومات IF فقط، وإنما تتوفر في العديد من النظم المعالجة للمعلومات النصية مثل نظم استرجاع المعلومات IR، ونظم الاستخلاص والتصنيف extraction and categorization systems. ويمكن ايضاح العلاقة فيما يلي:

- نظم التصنيف categorization systems: وتهدف هذه النظم إلى ربط وحدات البيانات بفئات محددة مسبقا. وتعد عملية التصنيف مشابه لعملية الترشيح، والتي تطابق فيها وحدات البيانات بسمات المستفيد. والسمة المميزة لنظم التصنيف أنها ثابتة بطبيعتها، حيث لا تتغير في كثير من الأحيان، في حين أن نظم الترشيح نظم ديناميكية حيث أن سمات المستفيدين قد تكون متغيرة.
- نظم التكشيف Extraction systems: تتعامل مع استخراج المعلومات من النصوص الواردة، وتتم عملية التكشيف من خلال تفعيل مجموعة من المعايير المحددة سلفاً. ولا تهتم هذه الأنظمة بالصلة الموضوعية للمعلومات المكشفة، بينما الصلة الموضوعية للمعلومات ذات درجة كبيرة من الأهمية في نظم ترشيح المعلومات. [٢٤]

بالرغم من فوائد الترشيح إلا أن هناك صعوبات في تحديد المعلومات التي يرغب في أن يطلع عليها المستفيد، فقد يتطلب توصيف المعلومات التي تمثل اهتمام المستفيد الرجوع إلى " سمات المستفيد User Profile "، ويمكن أن يتم التعرف بشكل إيجابي على سمات المستفيد من خلال تصميم نموذج لاهتمامات المستفيد. [٢٥]

ويرى الباحث ان جهود الباحثين في مفهوم الترشيح ركزت على علاقة الترشيح بنظم معالجة المعلومات مثل نظم الاسترجاع ونظم التكشيف ونظم التصنيف،

---

[23] Oard, Doug . Loc. Cit.
[24] URI HANANI, BRACHA SHAPIRA and PERETZ SHOVAL.Loc. cit. .- 205.
[25] Foltz, Peter W., Dumais, Susan T. Loc. cit

واتجاه المتخصصين في مجال الحاسبات لتطبيق مفهوم الترشيح في برمجيات تجارية تتعامل مع نظم استرجاع معلومات، كما ابرزت الأدبيات أهمية العائد المعلوماتي على المستفيد ومدى رضاه عن مخرجات نظام استرجاع المعلومات في ظل تطبيق آليات الترشيح، حيث يعد جوهر ترشيح المعلومات هو الحاجات المعلوماتية للمستفيد والتي يصفها بدقة في نظام الاسترجاع حتى تتوائم المخرجات مع احتياجاته المعلوماتية.

٢/٢/٢ أوجه الاتفاق والاختلاف بين الترشيح واسترجاع المعلومات

تحظى العلاقة بين الترشيح واسترجاع المعلومات باهتمام الباحثين من حيث أوجه الاتفاق و الأختلاف بينهما، لذا سيتناول الباحث بالنقاش هذه القضية لإبراز العلاقة بين الترشيح واسترجاع المعلومات من واقع المعالجة في ادبيات الانتاج الفكري.

يبدوا أن الفرق بين ترشيح المعلومات واسترجاع المعلومات يشوبه بعض التشويش. حيث تركز كلا من أنظمة الترشيح ونظم استرجاع المعلومات بشكل عام على اختيار المعلومات، كما يبنى كلا النظامين على نفس الأسس. إذا أن واحد من الاعتبارات العامة لاسترجاع المعلومات هي اختيار المعلومات، فيما يقوم ترشيح المعلومات باختيار المعلومات في بيئة الويب.[٢٦]

كما نجد أن العنصر المشترك بين كلا من استرجاع المعلومات وترشيح المعلومات، يتم في قاعدة البيانات النصية من خلال وصف مداخل الوثيقة بكلمات تمثل مصفوفة الوثيقة Document Matrix، التي تمثل تردد ورود الكلمة في الوثيقة. وهكذا، يمكن أن نفكر في الوثيقة ككمية موجهة Vectors في فضاء متعدد الأبعاد Multi-Dimensional Space|، ومن هذه الأبعاد الكلمات المستفيد في وصف النصوص. وفي معيار "مطابقة الكلمة المفتاحية Keyword Matching" بنظام التوجيه، فأن التشابة بين وثيقتين يحسب على أساس المنتج الداخلي Inner-Product أو جيب تمام Cosine المطابقة لعمودين من أعمدة الكلمات بمصفوفة الوثيقة. ويمكن

[26] Oard, Doug . Loc. Cit.

تمثيل الاستفسارات ككلمات موجهه وهكذا يتم مقارنتها في مقابل كل أعمدة الوثيقة ويتم استرجاع أفضل المطابقات.[٢٧]

ومن ناحية أخرى فإذا كان المفهوم الشخصي للمستخدم عن استرجاع المعلومات يتضمن اختيار المعلومات الساكنة نسبياً ذات العلاقة للإجابة على الاستفسارات الدينامية نسبياً، فمن المؤكد أن ترشيح المعلومات سيقدم رؤية أفضل لحل مشكلات لاسترجاع المعلومات. ومع ذلك فأن من الواضح أن الباحثون في ترشيح المعلومات من المحتمل أن يستفيدوا من الإنتاج الفكري المعروف لسمات استرجاع المعلومات.[٢٨]

ويختلف ترشيح المعلومات عن استرجاع المعلومات بمفهومه التقليدي في أن المستفيدين لهم اهتمامات موضوعية طويلة المدى يمكن توصيفها من خلال سمات المستفيد، بالاضافة إلى الاحتياجات الموضوعية الأخرى المتغيرة والتي يمكن رصدها من خلال الاستفسارات الموجهة لبعض نظم استرجاع المعلومات.[٢٩]

ويرى كلا من Robert WyattNasraoui OlfaZhuhadar Leyla, , الاختلاف الجوهري بين الترشيح واسترجاع المعلومات بمفهومه التقليدي يكمن في أن الترشيح يركز على التعامل مع المعلومات الديناميكية المتغيرة عكس نظم الاسترجاع التقليدية التي تتعامل مع المعلومات الساكنة مثل قواعد البيانات.[٣٠]

وقد عرضا كلا من Belkin and Croft نظام ترشيح المعلومات كنوع خاص من نظم استرجاع المعلومات، مما يشير إلى أن نظام ترشيح المعلومات سيتعامل مع ذات المشكلات الناتجة عن البحث والاسترجاع والتي تتعامل معها نظم استرجاع المعلومات. وتشترك كلا من نظم استرجاع المعلومات ونظم ترشيح

---

[27] Foltz, Peter W., Dumais, Susan T. Loc. Cit.

[28] Oard, Doug . Loc. Cit.

[29] Kuflik , Tsvi and Shoval, Peretz. Loc. Cit.

[30] Zhuhadar Leyla, Nasraoui Olfa, Robert Wyatt. Loc. Cit .- P. 1.

المعلومات في هدف عام واحد الا وهو اختيار المعلومات ذات الصلة باهتمامات المستفيدين، ولكنهما يختلفان في الخمس جوانب التالية:

- تكرار الاستخدام: صممت نظم استرجاع المعلومات بغرض أن يستخدمها المستفيد في كل مرة لتلبية احتياج معلوماتي وقتي ومحدد، في حين صممت نظم ترشيح المعلومات بغرض استخدام المستفيد لها على المدى الطويل بصفة متكررة لتلبية احتياجاته المعلوماتية طويلة المدى.
- **نموذج الاحتياجات المعلوماتية**: يتم تمثيل احتياجات المستفيد المعلوماتية والتعبير عنها في نظم استرجاع المعلومات على شكل استفسارات بحثية، في حين يتم تمثيل ووصف احتياجات المستفيد المعلوماتية بنظم ترشيح المعلومات في سمات المستفيد User Profiles نظراً للحاجة إلى استخدامها على المدى الطويل.
- **الهدف**: تختار نظم استرجاع المعلومات البيانات والوثائق ذات العلاقة والتي تطابق الاستفسار البحثي من قاعدة البيانات، بينما تحجب نظم ترشيح المعلومات البيانات والوثائق غير ذات العلاقة من نتائج الاسترجاع، أو تجمع وتوزع البيانات والوثائق ذات العلاقة المسترجعة من مصادر معينة بما يتوافق مع سمات المستفيد user's profile.
- **قواعد البيانات**: تتعامل نظم استرجاع المعلومات مع قواعد البيانات الثابتة نسبيا static databases مثل قواعد البيانات التي تتاح على CD-ROM، بينما تتعامل نظم ترشيح المعلومات مع قواعد البيانات الديناميكية dynamic data مثل رسائل البريد الإلكتروني.
- **أنواع المستفيدين**: تخدم نظم استرجاع المعلومات المستفيدين غير المعروفين للنظام؛ حيث يمكن لأي مستخدم الوصول للنظام وتقديم استفسار بحثي، بينما يجب ان يكون المستفيد بنظم ترشيح المعلومات معروف للنظام من خلال نموذج للمستخدم والذي عادة ما يكون مخزن في نموذج سمات المستفيد user's profile.

- **قضايا أخرى**: تهتم نظم ترشيح المعلومات بالقضايا الاجتماعية مثل نمذجة المستفيد user modeling والخصوصية والتي لا تقع ضمن اهتمامات نظم استرجاع المعلومات بدرجة كبيرة.[٣١]

بينما يرى كلا من Robert WyattNasraoui OlfaZhuhadar Leyla, , أنه يمكن تحديد الاختلافات بين الترشيح واسترجاع المعلومات فيما يلي:

- تهتم نظم الترشيح باهتمامات المستفيد البحثية طويلة الأجل، لذا فإن الترشيح بمثابة نظام ينشئ ملف تراكمي لأنشطة المستفيد، بينما استرجاع المعلومات بالاهتمامات البحثية الوقتية، لذا فإن النظام يتعامل مع تقديم المعلومات ذات الصلة بموضوع الاستفسار كمهمة واحدة محددة.

- يفترض في ترشيح المعلومات أن سمات المستفيد User profile محددة بدقة، لذا فإن المعلومات المقدمة تلائم إحتياجات المستفيد، على الجانب الأخر، فإن الأمر المؤكد في استرجاع المعلومات أن الوصول للمعلومات التي تلبي الاستفسار المعلوماتي للمستفيد أمر صعب لأنه يعتمد في الأساس على صياغة الاستفسار البحث بشكل دقيق للحصول على نتائج دقيقة.

- يعد نشر ووصول المعلومات للأفراد أو المجموعات بمثابة الاهتمام الرئيسي للترشيح، بينما يعد تجميع وتنظيم المعلومات بمثابة الاهتمام الرئيسي لاسترجاع المعلومات.

- يركز الترشيح على ترشيح وحدف المعلومات غير المرغوب فيها من قبل المستفيد، بينما يركز استرجاع المعلومات على الوصول للمعلومات ذات العلاقة بالاستفسار المعلوماتي. [٣٢]

ويمكن تلخيص الهدف الرئيس لترشح المعلومات كطريقة لاسترجاع المعلومات، في قدرته على التعامل مع مشكلة تقديم المعلومات ذات الصلة باهتمامات

[31] URI HANANI, BRACHA SHAPIRA and PERETZ SHOVAL. Loc. Cit .-P 204.
[32] Zhuhadar Leyla, Nasraoui Olfa, Robert Wyatt, Loc. Cit.- P.P 1-2.

المستفيد، وتحديدا المعلومات التي تتناسب وتفضيلات المستفيد، والتي يشار إليها بسمات المستفيد User profile. حيث يتعامل ترشيح المعلومات مع الحاجات المعلوماتية طويلة الأجل للمستفيد في مجال موضوعي محدد. [٣٣]

ويرى الباحث أنه يمكن إيجاز ما سبق في أن المستفيد يسترجع أو يتلقى المزيد من المعلومات بصفة دائمة ومنتظمة، ويستنفذ جزء كبير من وقته في تقييم هذه المعلومات، في حين أن كثير من المعلومات المتدفقة عليه بانتظام تكون دون المستوى الذي يرغبه. لذا يرى البعض أن برامج الترشيح هي الطريقة الوحيدة للحصول على أفضل محتوى يلائم احتياجات المستفيد. ويجب النظر إلى برامج الترشيح بصورة إيجابية، حيث أنها تساعد في اختيار المعلومات المناسبة وفي نفس الوقت تستبعد المعلومات غير المناسبة.[34]

ويوضح العرض السابق لأوجه الاتفاق والاختلاف بين الترشيح واسترجاع المعلومات في أدبيات الانتاج الفكري الصلة الموضوعية بين كلا المفهومين، حيث تم التعاطي مع المفهومين لإيضاح الصلة الوثيقة بين الترشيح كمفهوم وآلية تطبيق أكثر تحديداً والتصاقاً باهتمامات المستفيد البحثية طويلة المدى؛ وبين استرجاع المعلومات كمفهوم وآلية تطبيق تهدف لتلبية اهتمامات المستفيد البحثية الآنية، وغن كان نجاح النظام المبني على كلا المفهومين مرتبط بالدرجة الأولى على قدرة المستفيد على تحديد احتياجاته المعلوماتية بدقة.

لذا يخلص الباحث أن هناك علاقة وثيقة بين الترشيح واسترجاع المعلومات، حيث يعد الترشيح احد أدوات استرجاع المعلومات الهامة في بيئة الويب لتلبية احتياجات المستفيد في ظل التضخم المعلوماتي لنظم استرجاع المعلومات المفتوحة، وبالرغم من أنه قد يكون هناك بعض الاختلاف في التطبيقات والآليات ولكن مع تقارب المفهوم ووحدة الهدف وهو تلبية احتياجات المستفيد وتحقيق رضاه المعلوماتي،

[33] Op Cit.- P.2.

[34] Ben Gross. "Information Filtering". - Last update 03/12/2006 .- access at April 2006 . – URL: http://bengross.com/filter.html.

وتقود الباحث هذه النتيجة لإيضاح آليات الترشيح واسترجاع المعلومات في بيئة الويب.

٣/٢ آليات الترشيح وعلاقته باسترجاع المعلومات

مما لاشك فيه أن الآليات المطبقة في الترشيح تعكس وبوضوح العلاقة بين الترشيح واسترجاع المعلومات، ويرى الباحث أن تطبيق مفهوم الترشيح في بيئة الويب أفرز قناعات جديدة منها، أن شبكة الإنترنت وترشيح المعلومات يمثلان وجهان متضادان لعملة التقنية. فالإنترنت تزود المجتمع بالوسائل المختلفة للوصول إلى المعلومات. أما الترشيح فهو طريقة لمنع الوصول إلى المعلومات. ويتم تحميل برامج الترشيح مع متصفح الويب لتتولى مهمة منع الوصول إلى مواد على الإنترنت تصنف على أنها مواد غير ملائمة أو غير مناسبة لفئة معينة من المستفيدين، وتتفاوت الوسائل المستفيده للمنع بين برامج الترشيح. [٣٥]

تختلف الآليات المتبعة في الترشيح ولكنها تتفق في الهدف، ويتوقف تطبيق آلية الترشيح على احتياجات المستفيد المعلوماتية، لذا سيعرض الباحث بالعرض والتحليل لاليات الترشيح المختلفة لإبراز انعكاسها على استرجاع المعلومات وفق الحاجات المعلوماتية للمستفيد بهدف إيضاح العلاقة بين الترشيح واسترجاع المعلومات.

جذب ترشيح المعلومات الانتباه مؤخرا استنادا على دراسة استقصائية لتبادل المعلومات بين المنظمات، وقد توصلت الدراسة لثلاث طرق يعتمد عليها المستفيد في اختيار الوثائق، حيث يصنف ترشيح المعلومات إلى ترشيح معرفي Cognitive filtering system، وترشيح اجتماعي Social filtering system، وترشيح اقتصادي Economic filtering system. ويستند الترشيح المعرفي في اختيار الوثائق على محتواها، بينما يستند الترشيح الاجتماعي على توصيات المستفيدين

[35] Weaver, Lionel W. "Internet Filters, Public Libraries, and the First Amendment" . - access at April 2006 . – URL: http://www.uiowa.edu/~cyberlaw/cls00/weaverl.html

الأخرين بالوثيقة، أما الترشيح الاقتصادي فيستند على آلية لحساب التكلفة والعائد على المستفيد من استرجاع الوثيقة.[٣٦]

ويرى الباحث أن منهجيات الترشيح وآلياته تختلف باختلاف سمات المستفيدين في البحث والوصول للمعلومات، والاهتمامات البحثية المختلفة باختلاف مجال الاهتمام، فالترشيح للمعلومات المتعلقة بالاهتمامات الثقافية للمستفيد يختلف في المنهجية عن الترشيح في للمعلومات المرتبطة بالاهتمامات التخصصية والبحثية، ويؤكد هذا الاتجاه اختلاف التطبيق للمنهجيات حيث

١/٣/٢ طرق الترشيح

تتم عمليات الترشيح بأكثر من طريقة، تتمثل في:

- الترشيح الفردي individual filtering

ويتم من خلالها تحديد معلومات توضح اهتمامات كل شخص وما يفضله بصورة فردية.

- الترشيح المشترك أوالجماعي collaboration filtering

وتتم عبرها تنقية المعلومات لجماعة من المستفيدين، قد يكونون من منطقة إقليمية معينة.[٣٧]

- الترشيح القائم على المحتوى content-based - filtering

يعتمد على استرجاع الوثائق للمستفيد وفقا للعلاقة بين محتواها وتفضيلات المستفيد، كأن يتم مقابلة الكلمات المفتاحية في الوثيقة والكلمات المفتاحية المعبرة عن تفضيلات المستفيد.[٣٨] وتعتمد على مقارنة النص بسمات المستفيد، التي يتم تحديدها من

[36] Zhuhadar Leyla, Nasraoui Olfa, Robert Wyatt. Loc. Cit .- P.P 1-2.

[37] فاتن سعيد بامفلح. مصدر سابق. ص ٦٣

[38] Shardanand,Upendra and Pattie Maes. Social Information Filtering: Algorithms for Automating "Word of Mouth".- ACM Press/Addison-Wesley Publishing Co. New York, NY, USA ©1995
.-accessed at 20 October 2011.- Available at:
http://www.sigchi.org/chi95/proceedings/papers/us_bdy.htm

خلال الكلمات المفتاحية التي يستخدمها للتعبير عما يبحث عنه، كذلك يتم من خلالها تحليل المواد التي يتم تقييمها من قبل المستفيد، وإنشاء سمات المستفيد اعتمادا على وصف محتوى تلك المواد. وتستخدم السمات بعد ذلك للتنبؤ بتقديرات المواد التي يتم استعراضها مستقبلاً، والتي يتوقع أن تمثل اهتمامات مستقبلة للمستخدم. [٣٩]

- الترشيح التعاوني التقليدي traditional collaboration filtering

وتقوم على فكرة التعاون بين المستفيدين لمساعدة بعضهم في إنجاز الترشيح؛ حيث يقدم المستفيد تغذية راجعة تمثل ردة فعله تجاه المواد التي يستخدمها، وذلك في شكل تقييم rating وتستخدم تلك المعلومات في تقديم توصيات للمستخدمين ذوي الاهتمامات الموضوعية المشابهة. [٤٠] وتطبق هذه الآلية في خدمة GoodNews، كما يمكن أن يعتمد الترشيح التعاوني على معايير شخصية كتفضيلات صديق، او معايير جماعية كتفضيلات مجموعة، وهناك مواقع تعتمد على البيئة التعاونية في قبول المعلومات من مصادر عديدة، إلى جانب التفضيلات الشخصية. [٤١]

- الأسلوب المختلط Hybrid Techniques

وتتم وفقا لهذا الأسلوب المزاوجة بين الترشيح القائم على المحتوى، والترشيح التعاوني وفي سبيل إضفاء الطابع الشخصي على عمليات البحث والاسترجاع وتنقيتها، يتطلب الأمر من النظام جمع معلومات عن المستفيد، وذلك بغرض تحديد السمات الخاصة به، أو ما يعرف profile حيث يعتمد النظام على تلك السمات المحفوظة لتقديم معلومات تطابق الاحتياجات الفعلية المناسبة للسمات الشخصية للمستخدم. [٤٢]

حيث يطبق الترشيح المعرفي والتعاوني في اختيار الوثائق التي تهم المستفيد، ولكن يكمن الاختلاف بين منهجيتي الترشيح في مجال التطبيق، حيث يفضل احد

---

[39] فاتن سعيد بامفلح. مصدر سابق. ص ٦٣

[40] المصدر السابق. ص ٦٣

[41] Zhuhadar Leyla, Nasraoui Olfa, Robert Wyatt. Loc.Cit.- P.P 1-2.

[42] فاتن سعيد بامفلح. مصدر سابق. ص ٦٤

المناهج على الأخر في بعض الأحيان، ففي حال رغبة المستفيد الوصول للمعلومات بغرض متابعة كل ما هو جديد في مجتمع معلوماتي محدد سيفضل استخدام الترشيح التعاوني لاختيار الوثائق، أما في حال كانت رغبته الوصول للمعلومات في موضوع محدد سيفضل الترشيح المعرفي لأنه اكثر ملائمة لاحتياجاته المعلوماتية، واحيانا يستخد مزيج من المنهجيتين hybrid approach. [٤٣]

- سمات المستفيد User Profile

تعرف السمات بأنها تمثيل نظامي لاهتمامات الشخص، فهي بمثابة معلومات عن المستفيد تعبر عن احتياجاته من المعلومات . وقد كانت تستخدم فيما مضى بديلاً عن الاستفسار في نظم استرجاع المعلومات، أما في الوقت الحاضر فقد تم تحقيق التكامل بين السمات والاستفسارات في سبيل تحقيق درجة رضا أكبر عن النظام بين المستفيدين. [٤٤]

وقد ركز العديد من الباحثين على ابتكار آلية للتعرف والمحافظة على "سمات المستفيد user profiles " وتطبيق هذا الأسلوب في استرجاع المعلومات. وقد أكد منتجي البرامج على فكرة التطوير والتحسين لبرامج الترشيح بمرور الوقت وذلك بالاستعانة بالتغذية المرتدة للمستخدمين. [٤٥]

وقد أسهم تحديد السمات في تقديم خدمات المعلومات للمستفيدين من المكتبات ومراكز المعلومات قبل أن يستخدم لتحسين عملية استرجاع المعلومات من محركات البحث ونظم المعلومات الإلكترونية، وذلك من خلال خدمات الإحاطة الجارية، والبث الانتقائي للمعلومات، والخدمة المرجعية، التي تعمل كل منها على تحديد سمات المستفيدين. [٤٦]

وقد ظهرت أقدم أشكال ترشيح المعلومات الإلكترونية من العمل في البث الانتقائي للمعلومات Selective Dissemination of Information (SDI)، وهي

---

[43] Zhuhadar Leyla, Nasraoui Olfa, Robert Wyatt. Loc.Cit.- P.P 1-2.

[44] فاتن سعيد بامفلح. مصدر سابق. ص ٦٤

[45] Nabil R. Adam ...[others] . Loc. Cit.

[46] فاتن سعيد بامفلح. مصدر سابق . ص ٦٤

خدمة صممت كطريقة آلية لإبقاء العلماء على إطلاع مستمر بالوثائق الجديدة التي نشرت في تخصاصتهم. ويمكن للباحث أن يحدد أو يعدل الكلمات المفتاحية الخاصة به التي تصف مجال اهتمامه. لذا فأن البث الانتقائي للمعلومات يستعمل أسلوب المستفيد User Profile لمضاهاة الكلمات المفتاحية مع الوثائق الجديدة لتحديد أي هذه المقالات يقع في مجال اهتمام الباحث، وبالرغم من تطبيق خدمة البث الانتقائي على نطاق واسع فأن استعمال هذه الخدمة كان اقل بكثير من المتوقع. [٤٧]

وتنقسم سمات المستفيد إلى نوعين:

- **السمات البسيطة Simple Profile**

تشبه كثيرا الاستفسارات التي يقدمها المستفيد إلى نظم الاسترجاع؛ وتتكون من مجموعة من المصطلحات المفتاحية المحددة أوزانها في الغالب. وتستخدم عند تقديم خدمات معلومات مستفيدين ثابتين نسبياً، ولديهم اهتمامات دائمة؛ ومن ذلك خدمة الإحاطة الجارية، والبث الانتقائي للمعلومات، حيث يتم تسجيل بيانات تعبر عن اهتمامات المستفيدين في شكل مصطلحات، ويتم تحديثها كل فترة.

- **السمات الممتدة Extended Profile**

تتضمن معلومات أكثر صعوبة من تلك التي ترد في السمات البسيطة، وهي معلومات تتعلق بالشخص الذي يستخدم المعلومات بدلا من أن تتعلق باحتياجات محددة من المعلومات؛ فهي تساعد على تكوين خلفية من المعلومات عن المستفيد حتى تساعد في تحديد الوثائق التي يتم استرجاعها له. ومن أمثلة ذلك :جمع معلومات عن المستفيد من قبل اختصاصيي المراجع خلال المقابلة المرجعية،وتركز تلك المعلومات عادة على جوانب مختلفة من بينها :المستو ى التعليمي للمستفيد، واللغات التي يفضلها، والمجلات العلمية التي يشترك فيه، ومدى تآلفه مع الموضوع. [٤٨]

---

[47] Foltz, Peter W., Dumais, Susan T. Loc. Cit.

[48] فاتن سعيد بامفلح. مصدر سابق. ص ٦٤

ولكن هناك العديد من المشكلات في تطوير نموذج جيد لسمات المستفيد user profiles منها؛ استخدام طرق عديدة في توصيف مجالات اهتمام المستفيد. ويعد الأسلوب الشائع الاستخدام تقديم المستفيد مجموعة من الكلمات التي تصف مجالات اهتمامه، وهناك أيضاً العديد من الطرق الأخرى التي يمكن استخدامها لتحديد مجالات اهتمام المستفيد، مثل المقالات التي قرءاها المستفيد في الماضي، والمؤسسة التي يعمل بها، والكتب التي يطلبها المستفيد. [٤٩]

وإذا كانت عوامل تساعد في توقع ما يقع في مجال اهتمام المستفيد مثل تحديد القطاع الموضوعي العام لقراءاته أو محتوى المقالات، فإن هناك عوامل أخرى قد تكون مفيدة في توقع المعلومات التي يرغب المستفيد في الإطلاع عليها مثل الألفة، والجديد، والأهمية أو الحاجة الملحة، بالإضافة إلى أنه قد يكون هناك تفاعلات بين عدة عوامل في ذات الوقت. ولكن تجدر الإشارة إلى أن هناك أيضاً عوامل أخرى قد تكون غير مفيدة أو غير فعالة في وصف اهتمامات المستفيد البحثية مثل تقدم وصف لاهتمامات المستفيد بناء على تصفحه إلى أخبار العالم. ويجب التأكيد على أن هذه العوامل قد تساعد في توقع اهتمامات المستفيد، بالرغم من انه وليس هناك ضمان أن هذه العوامل يمكن تحديدها بسهولة. [٥٠]

ويلاحظ الباحث أن منهجيات الترشيح وإن اختلفت في التطبيق ولكنها تتفق في تحقيق الفاعلية في المخرجات المسترجعة وفي تلبية احتياجات المستفيد ووجهة نظره في تحديد احتياجاته وسلوكه المعلوماتي، فقد يفضل المستفيد الاسترجاع بناء على المحتوى على أن يتخذ قراره بتحقيق الوثيقة لاحتياجاته المعلوماتية من عدمه كما في الترشيح المعرفي، أو يميل المستفيد للترشيح الاجتماعي والذي يعتمد فيه على أراء الأخرين لتوفير الوقت وزيادة الموثوقية في نتائج الاسترجاع كما يطبق في الدرسات الببليومترية من حيث عدد الوثائق المستشهدة بوثيقة معينة؛ والتقديرات الممنوحة للصفحات والمواقع على شبكة الإنترنت، او تكون نتائج الاسترجاع من قواعد بيانات

[49] Foltz, Peter W., Dumais, Susan T. Loc. Cit

[50] Op. Cit

مرسمة أو دور نشر تستلزم حساب التكلفة من الاسترجاع؛ فيكون العامل الحاسم للمستفيد حساب العائد المعرفي كقيمة مضافة أو كمردود مادي من توظيف محتوى الوثيقة المسترجعة، وإن كان الباحث ان المستفيد يتعامل مع نتائج الاسترجاع بمفاهيم الترشيح الثلاثة حسب طبيعة النتائج المسترجعة ومدى احتياجه المعلوماتي لها.

وقد خلص الباحث إلى أن سمات المستفيد أساس نجاح نظام ترشيح المعلومات، لذا فإن بناء ملف سمات المستفيد يعد الخطوة الأولى للإنتقال من مرحلة استرجاع المعلومات التي تهدف لتلبية الاحتياجات المعلوماتية الأنية، إلى مرحلة الترشيح التي تهدف لتلبية الاحتياجات المعلوماتية الممتدة والقابلة للتعديل وفق تغير اهتمات المستفيد واحتياجاته البحثية، لذا سيسعى الباحث إلى إلقاء الضوء على آلية تمثيل سمات المستفيد بطريقة صحيحة تسهم في تحقق الهدف من ترشيح المعلومات.

يتم تمثيل سمات المستفيدين بأساليب متعددة، تتمثل في الآتي:

١. سمات الكلمات المفتاحية keyword profile

وهو الأسلوب الأسهل والأكثر شيوعا لبناء سمات المستفيد، ويعتمد على استخدام مجموعة من الكلمات المفتاحية التي يمكن أن يتم اقتباسها آلياً من الوثائق التي يزورها المستفيد أو من الإشارات المرجعية التي يتم حفظها لصفحات الويب bookmarked web pages، أو من الكلمات المفتاحية التي قدمها المستفيد صراحة.[٥١]

وتعد إحدى الطرق البسيطة لتحديد ما إذا كانت المعلومات تطابق اهتمامات المستفيد من خلال مطابقة الكلمات المفتاحية للوثيقة مع الكلمات المفتاحية المستخدمة في وصف اهتمامات المستفيد، وبالتالي فأن الوثائق أو مصادر المعلومات التي تحتوي على هذه الكلمات يجب أن تكون ذات علاقة. ويتم تجاهل التطابق الدقيق للكلمات المفتاحية مثل حالة الأحرف وغيرها، وقد تظهر نتائج غير مطابقة وغير ملائمة وذلك لأن الكلمات التي يستخدمها المستفيد لا تعكس بدقة الموضوع أو المحتوى بشكل

[51] فاتن سعيد بامفلح. مصدر سابق. ص ٦٥

واضح. فهناك مشكلات مثل المترادفات والتجانس اللفظي وغيرها من المشكلات التي تؤثر على عملية الاسترجاع.[٥٢]

وقد طبقت هذه الآلية في التكشيف الدلالي الكامن Latent Semantic Indexing (LSI) وموجهات مطابقة الكلمات المفتاحية Keyword Vector Matching حيث يتم تمثيل الوثائق كموجهات في فضاء واسع الأبعاد High Dimensional Space. وتتقرر في موجهات الكلمة المفتاحية KVM القيم في كل بُعد بناء على أي الكلمات موجودة في الوثيقة. أما في موجهات LSI فيتم تقرير القيم استناداً على أصغر رقم من أبعاد التكشيف المستنتجة بشكل إحصائي. وتميل الوثائق في مجالات موضوعية مماثلة إلى التقارب من بعضهم البعض بسبب اشتراكهم في الكلمات (في مطابقة الكلمة المفتاحية) أو قيم التكشيف (في LSI). وتستعمل هذه الميزة كأساس للترشيح.

وبوجه عام، فإن فكرة الترشيح هي إنشاء فضاء الوثائق Space Of Documents، وهي الوثائق التي سبق عليها الحكم من قبل المستفيد بأنها تقع أو تتصل بمجالات اهتمامه. وإذا كانت هناك وثيقة جديدة قريبة الصلة من الوثائق ذات العلاقة في فضاء الوثائق، فمن المحتمل أن تكون مثيرة لاهتمام المستفيد. وبالمقابل، إذا كانت تلك الوثيقة بعيدة عن الوثائق ذات العلاقة، فمن ثم ستعتبر غير مثيرة لاهتمام المستفيد.

ويمكن أن يستخدم هذا الإجراء نفسه لتحديد مدى قرب أي وثيقة جديدة من اهتمامات المستفيد من خلال الكلمات المفتاحية في سمات المستفيدين User's Profile. وبعد كل هذه المقارنات، فأن الفرق الوحيد بين LSI و KVM هو أن LSI يمثل المصطلحات والوثائق في فضاء بعدي مصغر من أبعاد التكشيف الاشتقاقي.

[52] Foltz, Peter W., Dumais, Susan T. Loc. Cit

وقد قارن Foltz بين LSI وKVM في ترشيح مقالات الشبكة الإخبارية Netnews Articles من خلال إحدى التجارب حيث تم وضع التقدير الموضوعي لمقالات الشبكة الإخبارية على أساس إما أنها ذات علاقة أو ليست ذات علاقة باهتمامات المستفيد. التقديرات المبدئية لـــ ٨٠ % من المقالات التي قرأت ثبت أنها استعملت لتوقع المقالات ذات العلاقة والتي تمثل ٢٠% لكل شخص.وقد وجد Foltz أن ترشيح LSI يحسن مستوى التوقع للتفضيلات أكثر من طريقة KVM بمعدل ١٣% كما لوحظ أن هناك تحسن بمعدل ٢٦% في دقة تقديم المقالات في الطلبات المستلمة .[٥٣]

ويرى الباحث أن الهدف الرئيسي من استعمال كلا الطريقتين هو استبعاد الوثائق غير ذات العلاقة باهتمامات المستفيد، أما الوسيلة فتختلف فهي أما عن طريق الكلمات المفتاحية الدالة في الوثيقة كما في أسلوب مطابقة الكلمات المفتاحية KVM أو عن طريق مدى التقارب الموضوعي للوثائق كما في LSI، ولكن – من وجهة نظر الباحث - أن الطريقتان يشوبهما بعض المشكلات، فالطريقة الأولى KVM قد تكون ناجحة في حالة واحدة وهي الوصف الدقيق للكلمات الدالة في الوثيقة واستخدام نفس الواصفات في وصف اهتمامات المستفيد أي استخدام لغة مقيدة في الوصف والبحث. أما الطريقة الثانية LSI فليس من الضروري أن كل الوثائق ذات الصلة بوثائق يهتم بها المستفيد تكون أيضاً ذات علاقة باهتماماته. كما أن الترشيح بالمفهوم والطرق السابقة هو طريقة تفاعلية بين المستفيد ونظام استرجاع المعلومات.

٢. سمات الشبكة الدلالية semantic network profile

يتم استخدام الكلمات كنقاط nodes في الشبكة الدلالية .ولا يقتصر الأمر على استخدام الكلمات الفردية التي قد لا تكون دقيقة بدرجة كافية أحياناً، بـل يـتم تجميـع

[53] Op. Cit.

الكلمات ذات الصلة مع بعضها في مفاهيم يطلق عليها "المجموعات المترادفة synonym sets أو synsets" وبذلك تكون النقاط في شبكة الدلالات هي عبارة عن المجموعات المترادفة المرتبطة مع بعضها بعلاقات.

٣. سمات المفاهيم concept profile

وتشبه السمات القائمة على شبكة الدلالات، من حيث إن كلا منها يتم تمثيله بنقاط المفاهيم وعلاقات بين النقاط، أما في السمات القائمة على المفاهيم، فالنقاط تمثل موضوعات مستخلصة abstract topic تمثل اهتمامات المستفيد، وذلك بدلاً من كلمات محددة أو مجموعة من الكلمات المتصلة ببعضها) فالنظام هنا لا يتعامل مع الكلمات كمصطلحات في حد ذاتها، ولكن مع الموضوعات كمفاهيم . (كذلك فإن سمات المفاهيم تشبه نظيرتها الخاصة بالكلمات المفتاحية من حيث إنه في الغالب يتم تمثيلها بحيز من الخصائص الموزونة، ولكن الخصائص في هذه الحالة عبارة عن مفاهيم بدلاً من الكلمات .ويمكن بناء المفاهيم في شكل بسيط flat أو هرمي ، وهذا الأخير قد يتم بطريقة بسيطة اعتماداً على المكانز thesaurus أو قواميس المصطلحات reference taxonomy، وقد يتم بطريقة أكثر تعقيدا باستخدام خرائط المعرفة antologies. [٥٤]

ويرى الباحث أن تمثيل سمات المستفيد بالكلمات المفتاحية أو بسمات الشبكة الدلالية أو بسمات المفاهيم تعتمد بدرجة كبيرة على رصد الاهتمامات المعلوماتية الفعلية طويلة الأمد للمستفيد بدقة ليتم تمثيلها بالطريقة التي تتوافق مع نظام الاسترجاع لتحقق أعلى نسبة تحقيق في المخرجات المستدعاة بما يلبي احتياجات المستفيد المعلوماتية.

٢/٣/٢ تطبيقات الترشيح في الاسترجاع

لتأصيل أي مفهوم من الضروري رصد تطبيقه في المجتمع، وقد برزت نماذج لتطبيق الترشيح في مجال الاسترجاع على المستوى الأكاديمي والمستوى

[54] فاتن سعيد بامفلح. مصدر سابق. ص ٦٥

التجاري من خلال توظيف تقنيات الحاسب وتطوراته في تطبيق المفهوم، وسيعرض الباحث في هذا الجزء لعدد من تجارب التطبيق بهدف إبراز العلاقة بين الترشيح واسترجاع المعلومات.

فقد دخلت الأسواق الخدمات التجارية للترشيح، من خلال الخدمات الاخبارية التي ترسل بشكل شخصي للأفراد، وفق سماتهم الشخصية User profile وذلك باستخدام خدمات البريد الالكتروني والرسائل القصيرة، حيث يتم الترشيح باستخدام Smart system تحت إشراف بشري. [٥٥] حيث يجمع مجال ترشيح المعلومات بين أدوات من تقنيات الذكاء الاصطناعي AI (Artificial Intelligence) مثل الوكيل الذكي Intelligent Agent أو برمجيات الروبوت software robots ("softbots") مسترشدة بسمات المستفيد وأساليب استرجاع المعلومات بهدف تكشيف واسترجاع المحتوى. ويعتمد تطبيق نظم ترشيح المعلومات لتقنية الوكيل الذكي على توفير آلية للتجميع الآلي للمعلومات عبر شبكة الإنترنت أو المستودعات المعلوماتية مثل المكتبات الرقمية استناداً على ملف سمات المستفيد الذي يمثل احتياجاته المعلوماتية الأساسية، مما ينعكس على فاعلية وأداء نظم استرجاع المعلومات في تلبية الاحتياجات المعلوماتية للمستفيد. [٥٦]

وقد طورت مؤخرا العديد من الدراسات والنظم لاختبار أفكار ترشيح المعلومات، وقد أجرى باحث يدعى "Allen" سلسلة من التجارب لاكتشاف نماذج المستفيد في توقع التفضيلات للمقالات الإخبارية. في أحد التجارب توقع الباحث أي المقالات سيقوم الشخص بقراءتها مستنداً على المقالات السابقة التي قراءها نفس الشخص، مستخدما مقياس لتحديد التداخلات بين عناوين المقالات القديمة والجديدة. بينما كانت أفضل احتمالات التنبؤات عندما كان متوسط الارتباط بين المقالات المتوقعة وتقديرات موضوعات المقالات منخفض جداً بقيمة (r =0.44).

[55] Zhuhadar Leyla, Nasraoui Olfa, Robert Wyatt, Loc. Cit.- P.P 1-2.

[56] Kuflik , Tsvi and Shoval, Peretz. Loc. Cit

كانت أكثر النماذج نجاحا في توقع تفضيلات المستفيد للفئات العامة من المقالات عن المقالات المتخصصة. ولكن المهمة الصعبة هي توقع ما هي المقالات الإخبارية التي سيقرأها المستفيد، فموضوعات الأخبار تتفاوت من يوم لآخر مما يمثل صعوبة في الحصول على توقعات مستقرة عن اهتمامات الشخص، بالإضافة إلى أن المصادر الخارجية للأخبار من المحتمل أن تؤثر على ما يقراء الناس في التجربة، مع الاعتقاد بأن اهتمامات المستفيدين تجاه الإنتاج الفكري التقني سيكون أكثر استقراراً بمرور الوقت.

وقد استخدمت في بحث Allen التفضيلات السابقة للموضوعات في بناء نموذج ضمني لاسترجاع المواد ذات الصلة. ويمكن التعبير عن ذلك بطريقة مختلفة أن يترك للمستخدم حرية إيضاح بنية وتركيب المعلومات. ويسمح نظام عدسة المعلومات Information Lens system للمستعملين بابتكار القواعد لترشيح رسائل البريد مستندا على مطابقة الكلمة المفتاحية في حقول البريد. وبدأ ذلك منذ أن أصبح هناك تركيب بنائي محدد في رسائل البريد، مثل معلومات المرسل في حقل المرسل وكلمات مفتاحية في حقل الكلمة المفتاحية، ويمكن أن تستغل هذه القواعد في هذا التركيب ألبنائي لأداء ما يرغبه المستفيد من أعمال محددة على الرسائل. [٥٧]

وأثبتت النتائج أن الأشخاص الذين ليس لهم خبرة كبيرة في مجال الحاسب قادرين على ابتكار قواعد عدسة معلوماتهم Information Lens rules الخاصة لتفضيل وترشيح رسائلهم البريدية. وقد ابتكرت النسبة الأكبر من القواعد لمطابقة معلومات المرسل والمستلمين الأخريين، بينما ابُتكرت قواعد قليلة لمطابقة المعلومات النصية textual information مثل حقل الموضوع subject ونص الرسالة text body.

وبالرغم من تطوير مجموعة متنوعة من أنظمة المعلومات، فمازال هناك تقييم لهذه الأنظمة لتحديد أكثرها فاعلية في عملية الترشيح. ولكن مازال هناك العديد من الأسئلة بدون إجابات منها: ما هي الطرق الأكثر فاعلية لمطابقة اهتمامات المستفيد

[57] Foltz, Peter W., Dumais, Susan T. Loc. Cit.

والمعلومات المتوفرة؟، كيف يجب توصيف اهتمامات المستفيد؟، وكيف ستتفاوت تفضيلات طرق الترشيح في المجالات domains المختلفة .[٥٨]

يرى الباحث من العرض السابق أن لترشيح المعلومات دور في تحقيق الإفادة من مصادر المعلومات الضخمة والمتنوعة المتاحة على شبكة الإنترنت، وذلك من خلال ترشيح المعلومات وفقاً لاحتياجات المستفيد فيما يشبه خدمة البث الانتقائي للمعلومات في بيئة مصادر المعلومات التقليدية، بغرض تحقيق هدف واحد وهو تلبية احتياجات المستفيد من البيئة الرقمية لشبكة الإنترنت. كما أن أهم عنصر مؤثر في تطبيق آلية ترشيح المعلومات هو تلبية احتياجات المستفيدين، ويتم ذلك من خلال محاولة رسم تصور لاهتمامات المستفيد من خلال اطلاعاته السابقة User profile، وإن كانت هذه الطريقة قد يشوبها بعض العيوب نتيجة لتغير الاتجاهات القرائية والبحثية عند المستفيدين من وقت لآخر نظراً لتغير اهتماماتهم أما نحو درجة أدق من التخصيص أو نحو مجال موضوعي أخر، كما أنه ليس هناك طريقة معيارية لوصف احتياجات المستفيد، لذا لابد أولا من بناء نموذج لتوصيف اهتمامات المستفيد مع إمكانية تعديله وتحديثه مستقبلا في برنامج الترشيح ليتواكب مع احتياجات المستفيد المتغيرة.

٤/٢ الترشيح ونظم استرجاع المعلومات المفتوحة

تمثل محركات البحث أحد نظم استرجاع المعلومات المفتوحة باعتبارها نظم استرجاع مبنية على قواعد بيانات والتي تقوم بتجميع المعلومات بشكل آلي وحفظها وتحديثها واسترجاعها، حيث تتلقى قاعدة البيانات مدخلات متمثلة في مصادر المعلومات المتاحة في بيئة الويب، ليتم معالجتها وتقدم مخرجات بناء على الاستفسارات البحثية المقدمة من المستفيدين، وتتم عملية الجمع والحفظ بصفة مستمرة بواسطة برمجيات تتولى الحصر والإضافة لقاعدة بيانات النظام نظراً للنمو المتزايد لمصادر المعلومات في بيئة الويب؛ كما تقوم هذ البرمجيات بحذف مصادر المعلومات التي انتهت استضافتها في بيئة الويب من قاعدة بيانات النظام واستبعادها، لذا سيتناول

[58] Op. Cit

الباحث في هذا الجزء علاقة الترشيح بنظم الاسترجاع المفتوحة بهدف إبراز العلاقة بين الترشيح ونظم استرجاع المعلومات

منذ البداية الأولى للنظم الآلية لاسترجاع المعلومات في منتصف القرن العشرين، وظفت الخوارزميات واللوغريتمات الرياضية والمنطقية في الترتيب الطبقي لنتائج نظم استرجاع المعلومات، فتوالت الاسهامات في تطوير نماذج استرجاع المعلومات، كالنموذج البوليني Boolean retrieval، ونموذج فراغ الموجهات vector space model، ونموذج الاحتمالات probabilistic model، إلى أن سرعان ما تعطلت هذه الخوارزميات بسبب بيئة عمل جديدة تمثلت في الشبكة الويب، فأقدم المعلوماتيون على تطوير مجموعة من نماذج الاسترجاع وخورازميات الترتيب التي تلائم طبيعة الاسترجاع المحتوى في الشبكة الويب، كخوارزميات الترتيب المعتمدة على الروابط مثل نموذج ترتيب الصفحات PageRank ونموذج ترتيب المعتمد على الموضوع (Hypertext Induced Topic Search) HITS، إلى ان بعض محركات البحث قد نحنى منحنى اخر، حيث اعتمدت بعض محركات البحث على دمج خوارزميات الترتيب التقليدية مع خوارزميات الترتيب الخاصة بالويب واستخدامها لاسترجاع وترتيب المحتوى.[59]

١/٤/٢ نماذج استرجاع المعلومات في نظم المعلومات المفتوحة

تعتمد منطقية ترتيب الوثائق المسترجعة على حساب درجة التشابه بين الاستفسار والوثائق المكشفة، وعليه كلما ارتفعت درجة التشابه ارتفعت رتبة الوثائق المشابه للاستفسار. تاتي خوارزميات ترتيب المحتوى في محركات البحث على صعيد بسيط ومعقد على النحو الآتي:

[59] مؤمن سيد النشرتي. التحديات التي تواجه خوارزميات محركات البحث في استرجاع المحتوى العربي على الشبكة العنكبوتية العالمية: دراسة مسحية تحليلية.- ورقة بحثية مقدمة لمؤتمر المحتوى العربي على الانترنت.- ٣-٥ اكتوبر ٢٠١١ جامعة الامام محمد بن سعود الاسلامية. ص ١٨

- **النماذج البسيطة**

١. التحليل من خلال النقر: تعتمد هذه المنهجية على استخدام البيانات المتعلقة بتردد اختيار المستفيد لمحتوى صفحة بعينها استجابة للاستفساره كوسيلة للترتيب الطبقي او بمعنى اخر انها تعتمد على تسجيل استفسارات المستفيدين ومحددات المصادر الخاصة بمحتوى الوثائق المسترجعة، والتي قام بالنقر عليها للدلالة على مطابقتها للاستفسار المدخل، ومن ثم يحتل محتوى هذه الوثائق طبقة عليا عن غيرها في حالة إدخال نفس الاستفسار إلى محرك البحث.

٢. تحليل الروابط: تعد مرجعية هذه المنهجية إلى علم المعلومات والمكتبات حيث عُني بدراسة وتحليل الاستشهادات المرجعية، فمنهجية الروابط تعتمد على فحص الروابط التي تشير إلى محتوى الوثيقة، فمن خلالها تحدد درجة أهمية محتوى الوثيقة في موضوع محدد اعتماد على كم ونوعية الروابط التي تشير إلى هذا المحتوى.

٣. تردد المصطلح: وهي منهجية حسابية تعتمد على تقييم حساب تردد ظهور المصطلح في محتوى الوثيقة، فبشكل عام يدل التردد المرتفع لظهور الكلمات في محتوى الوثيقة على إحتمالية ان هذا المحتوى أشد ارتباطا بالاستفسار ويصحب هذه التقنية وجود ما يعرف بقائمة الاستبعاد.

٤. موقع المصطلح: في كثير من الاحيان يشير موقع المصطلح إلى أهميته في محتوى الوثيقة، ومن ثم أهمية محتوى الوثيقة في المجال الموضوعي التي تنتمي اليها، علاوة على ذلك تعتمد معظم محركات البحث إلى إعطاء وزن أكثر للمصطلحات التي تظهر بشكل جلي في اجزاء معينة من محتوى الوثيقة، مثل العنوان والفقرة الرئيسية والمستخلصات وتعليقات الصور عن نظيرتها التي تظهر في جسد الوثيقة أو في الهوامش السفلية.

٥. تقارب المصطلح: منطقية هذا العامل تتمثل في أن قرب المسافة بين مفردتين أو أكثر في الاستفسار يحقق نسبة مرتفعة في استرجاع محتوى الوثائق الذي يتشابه في قرب مفرداته مع الاستفسار، وهذا النمط يفضل في البحث عن أسماء الاشخاص والكيانات المعرفة.

٦. تنسيق النص: وهو نمط شكلي معني بتنسيق محتوى الوثيقة حيث أن الكلمات التي تتسم ببنط مختلف في كتابتها عن غيرها كجعلها بخط سميك تمثل أهمية في حمل محتوى الوثيقة إلى الطبقة الاعلى في النتائج المسترجعة.
٧. حقل العنوان: فبتدوين حقل العنوان في تكويد الوثيقة يوفر لها درجة مرتفعة في الترتيب الطبقي.[60]

ويرى الباحث أن النماذج البسيطة في ترتيب النتائج المسترجعة التي يتبعها محرك البحث تمثل احد اهم عيوب نظام استرجاع المعلومات بمفهومة التقليدي والمتمثلة في التعامل مع الاستفسار كمهمة واحدة منتهية، هذا إلى جانب أن المعايير التي يتم عليها تحديد علاقة الوثيقة باهتمام المستفيد مبنية على فرضيات التردد المتكرر دون النظر لزمن التصفح، أو فرضية تحليل الروابط التي من المحتمل أن تكون غير ذات علاقة وثيقة بالمحتوى، أو فرضية تردد المصطلح أو تقارب المصطلح مع الاستفسار أو تنسيق الكلمة في النص، قد تصيب هذه الفرضيات في استدعاء نتائج تلبي احتياجات المستفيد ولكنها في الوقت ذاته ستحقق معدل من الشوشرة في نتائج الاسترجاع تؤثر على رضا المستفيد.

- **النماذج المعقدة**

١. النموذج البوليني The Boolean model .

يأتي النموذج البوليني على رأس نماذج استرجاع وترتيب المحتوى في محركات البحث وأقدمها واكثرها خضوعا للدراسة، ويعتمد النموذج البوليني في أساسه على مفهوم نظرية المجموعات والجبر البوليني، ومنطقية عمله تعتمد على إقران محتوى الوثيقة بمجموعة من الكلمات المفتاحية لتمثيلها، وبالنسبة للاستفسار فيتم أيضا رصد الكلمات المفتاحية به وتحديد سمة الاقتران بين الكلمات المفتاحية من خلال الروابط البولينية. يعمل النموذج البوليني على توفير نظام يتمتع بالسهولة في ان يدرك

[60] المصدر السابق. ص ١٩

من قبل المستفيدين والمتعاملين مع محركات البحث علاوة على ذلك فإن استفسارات المستفيدين تحدد وفقا للروابط البولينية والتي تتسم بالدلالة المحكمة.[61]

ويرى الباحث ان النموذج البوليني يمكن تطبيقه في نظم الترشيح في بناء ملف سمات المستفيد لإبراز العلاقات بين الكلمات المفتاحية.

٢. النموذج ذات الموجهات في الفراغ Vector space model.

يعد نموذج الموجهات أحد وأهم نماذج الترتيب في استرجاع المحتوى استخداما وتوظيفا، وقد كان لبيتر لون الفضل في تطوير منطقية الاعتماد على الموجهات، حيث كان أول من اقترح استخدام النهج الاحصائي في البحث عن المعلومات عام ١٩٥٧ معتمدا على معيار التماثل بين الاستفسارات والوثائق. فوفقا للنموذج الاحصائي ينظر إلى محتوى الوثيقة على انه حقيبة كلمات Bag of words بمعنى أن محتوى الوثيقة يشتمل على مصطلحات غير مرتبة وذات ترددية غير منتظمة داخل محتوى الوثيقة.

كانت الفكرة الرئيسية التي بنيت عليها نموذج الموجهات في الفراغ أن استخدام الوزن الثنائي (يتصل أو لا يتصل (١،٠)) يحد جدا من عملية الاسترجاع والترتيب الطبقي للنتائج، وبناءا عليه قدم هذا النموذج اطار عمل جديدا يعتمد على ما يعرف بالمطابقة الجزئية (اي ان درجة اتصال أو عدم اتصال الوثيقة بالاستفسار يحدد من خلال أوزان متفاوتة بين قيمتين أدناها يرمز له بالرقم 0 وأعلاها يرمز له بالرقم 1 ويتم ذلك من خلال مجموعة من المعادلات الخاصة بوزن المصطلح بحيث تستخدم هذه الاوزان في نهاية المطاف لحساب درجة التشابه والتماثل بين كل من الوثيقة المختزنة في النظام وبين استفسار المستفيد).[62]

ويرى الباحث أن نموذج الموجهات يمكن تطبيقه في ترشيح المعلومات في بناء ملف سمات المستفيد في تحديد الموضوعات الدقيقة ذات الصلة باهتمامات

[61] المصدر السابق. ص ٢١
[62] المصدر السابق. ص ٢٢

المستفيد، وفي ذات الوقت استبعاد الموضوعات ذات الصلة ولكن تقع خارج دائرة اهتمام المستفيد.

٣. نموذج المنهج الإحتمالي Probabilistic model

قدم المنهج الاحتمالي لإول مرة في مضمار استرجاع المعلومات عام ١٩٦٠ على يد Maron and Kuhns في مقالتهما المعروفة بأسم On Relevance, Probabilistic Indexing and Information Retrieval كأول عمل علمي يتطرق إلى استخدام المنهج الاحتمالي في استرجاع المعلومات، وعليه ظهر ما يعرف بالتكشيف الاحتمالي Probabilistic Indexing.أما نموذج الاسترجاع الخاص بهذا المنهج فقد قدم في عام ١٩٧٦ على يد كلا من S. E. Robertson and K. Sparck Jones والذي عرف بأسم نموذج الاسترجاع القائم على الفصل الثنائي The binary independence retrieval model **(BIR)**، وتمثل مفهوم هذا النموذج في أن مجموعة الوثائق المختزنة في نظام استرجاع المعلومات تنقسم إلى مجموعتين ثنائيتين مستقليتين عن بعضهما البعض، المجموعة الاولى تعرف بمجموعة الصلة والتي يتسم محتواها بالصلة بالاستفسار، والمجموعة الاخرى تعرف بمجموعة اللاصلة والتي يتسم محتواها بعدم الصلة بالاستفسار. ويمثل جوهر هذا النموذج في سؤال منطقي وهو "**ماهو احتمالية صلة وثيقة محددة باستفسار محدد؟!**" من خلال هذا السؤال تبلورت رؤية هذا النموذج في قياس وتحديد الوثائق وفقا لإحتمالية صلتها بالاستفسار.[63]

يرى الباحث صعوبة تطبيق نموذج المنهج الاحتمالي في ترشيح المعلومات وبناء ملف سمات المستفيد، نظراً لاحتمالية استدعاء وثائق لا تقع في دائرة اهتمام المستفيد البحثي بما يتنافى مع جوهر فكرة الترشيح المتمثلة في استرجاع ما يتلائم مع احتياجات المستفيد.

٤. نموذج التكشيف الدلالي الكامن Latent Semantic Indexing Model

[63] المصدر السابق. ص ٢٥

إن نماذج الاسترجاع - السابق ذكرها - تعتمد وبشكل اساسي على الاسترجاع من خلال اختزال محتوى كلا من الاستفسار والوثيقة في مجموعة من المصطلحات الكشفية أو الكلمات المفتاحية. ثم يعمد بعد ذلك إلى قياس درجة التشابه بين كلا منهما، ثم الاسترجاع وفقا لهذا الاساس، وتعرف هذه المنهجية بإسم الاسترجاع وفقا للتشابه المعجمي lexical matching method أي من خلال الاعتماد على المضاهاة بين الاحرف المكونة للمصطلحات الكشفية الوارة في كلا من الوثيقة والاستفسار.[64]

عرفت هذه المنهجية بأسم التكشيف الدلالي الكامن Latent semantic indexing وقد قدم هذا النموذج كمهنجية آلية لمعالجة بعض أوجه القصور المتأصلة في تقنيات التي تعتمد على المضاهاة المعجمية والاسترجاع وفقا للكلمات المفتاحية في الوثيقة، يعتمد النموذج الدلالي الكامن على تحليل درجة العلاقة الدلالية بين محتوى الوثائق من خلال معدلات احصائية، فوفقا لهذه المنهجية فأن محرك البحث يقوم باسترجاع محتوى الوثائق ذات الصلة بمفهوم الاستفسار وليس وفقا لتشابه الكلمات المفتاحية حتى ولو لم تشتمل هذه الوثائق على مصطلحات الاستفسار.[65]

ويرى الباحث أن نموذج التكشيف الدلالي الكامن قد طبق فعليا في ترشيح المعلومات كما اوضحنا سابقاً في البحث، مما يبرز العلاقة بين الترشيح واسترجاع المعلومات.

٥. نموذج الاسترجاع (الروابط الفائقة الناجمة عن البحث الموضوعي (hyperlink-induced topic search) (HITS).

طورت هذه الخوارزمية عام ١٩٩٧ على يد Jon Kleinberg في نفس الوقت الذي طورت فيه خوارزمية الترتيب الطبقي، تعتمد هذه الخوارزمية على اكتشاف وترتيب محتوى الوثيقة ذات الصلة بموضوع محدد – وتعد هذه الخوارزمية الآن جزءا اساسي في خوارزميات محرك البحث Ask (www.ask.com) – بمعنى

[64] المصدر السابق. ص ص ٢٥-٢٦
[65] المصدر السابق. ص ١٥

ان هذه الخوارزمية تعتمد على أن يوجه المستفيد اولا الاستفسار لإداة البحث ثم تسترجع النتائج من الكشاف او قاعدة البيانات لتبدء مرحلة الترتيب للنتائج وفقا لعنصرين اساسين هما:

- المواقع الارتكازية Hubs Nodes: وهي المواقع التي تشتمل على محتوى يصدر منه الرابط في اشارة منها للمواقع الاستنادية.
- المواقع الاستنادية او ذات الموثقية Authorities Nodes: ويقصد بها المواقع التي يرد اليها الرابط من قبل المواقع المحورية.[66]

ويرى الباحث ان نموذج استرجاع HITS يصلح للتطبيق في ترشيح المعلومات نظراً لاعتماده على معيار الموثوقية للمواقع مما يحقق درجة من الثقة والأمان للمستفيد في النتائج المسترجعة بما يمثل احد اهداف ترشيح المعلومات.

٦. نموذج الاسترجاع (ترتيب الطبقي للصفحات PageRank).

طور هذا النموذج على يد كلا من Sergey Brin و Lawrence Page عام ١٩٩٧، وقد عرف هذا النموذج " بانها المنهجية التي تعني بحساب رتبة محتوى كل صفحة على الويب اعتمادا على نمذجة الويب في مخطط بياني قائم على الروابط والمواقع". ولقياس جدوى هذه الخوارزمية قاما كلا من Brin و Page بتصميم محرك البحث الشهير Google.

أن الجانب الذي التفتت اليه هذه الخوارزمية هو النظر إلى الكيف دون الكم، بمعنى الاخذ في الاعتبار جودة الروابط بدلا من النظر إلى عدد الروابط ، فتستند هذه الخوارزمية على مبدئين اساسين هما:

- تمثل الروابط مؤشرات جيدة لتحديد اهمية محتوى الوثيقة التي تشير اليها.
- الروابط الصادرة من وثائق تحظى باهمية في موضوعها تعد مؤشرا جيدا لجودة الوثيقة التي تشير اليها، عن الوثيقة التي يشار اليها من قبل وثائق اقل في الاهمية والجودة.[67]

[66] المصدر السابق. ص ١٩

ويرى الباحث أن نموذج الترتيب الطبقي للصفحات يصلح تطبيقه في ترشيح المعلومات، نظراً لتركيزه على وضع قيم للوثيقة استنادا على الروابط والتي يقابلها تحليل الاستشهادات المرجعية في علم المكتبات والمعلومات كمنهجية لترتيب الانتاج الفكري، مما يقدم مؤشرات تحليلية للمستفيد يمكن الاستناد عليها في المضاهاه مع ملف سمات المستفيد مما يحقق أعلى فائدة في الاسترجاع.

ويخلص الباحث من العرض السابق أن النماذج المعقدة في في ترتيب النتائج المسترجعة التي يتبعها محرك البحث - كنظام استرجاع معلومات مفتوح- منها المطبق في ترشيح المعلومات كنموذج LSI، كما يمكن تطبيق باقي النماذج في بناء ملف سمات المستفيد، ويعكس التقارب في التطبيق بوضوح العلاقة الوثيقة بين الترشيح واسترجاع المعلومات على مستوى آليات التطبيق مما يدعم الفرضية التي اطلقها الباحث.

٥/٢ الترشيح وتقييم نظم استرجاع المعلومات

يتفق الترشيح واسترجاع المعلومات في الهدف العام ألا وهو تلبية احتياجات المستفيد المعلوماتية، كما يعد الترشيح في نظم الاسترجاع المفتوحة احد آليات الاسترجاع التي تسهم في تحقيق هذا الهدف، لذا سيسعى الباحث لإبراز دور الترشيح كاحد عناصر تقييم أداء وكفاءة نظام الاسترجاع.

إن تصميم وتطوير نظام لاسترجاع المعلومات، يرتبط أساسا بتحقيق هدف أو مجموعة أهداف محددة، والتي ينظر إليها على إنها غاية النظام، وفي البدايات الأولى لنظم استرجاع المعلومات المحوسبة كانت قدرة النظام على استرجاع المعلومات هي هدفه الأول وغايته الأهم. ولكن بعد مرور مدة قصيرة من الزمن تبين أن استرجاع المعلومات كعملية مجردة وغير موجهه لا يمكن أن تكون هدفاً أوحداً للنظام، طالما إنها لا ترتبط بحاجة موضوعية أو بحثية لمستفيد ما، لذا فان حكم المستفيد بصلاحية المعلومات المسترجعة من النظام استجابة لمطلبه البحثي أصبحت

[67] المصدر السابق. ص ص ٢٨- ٢٩

في مقدمة أهداف نظم استرجاع المعلومات، وترتبط بها مقاييس تقييم نظم استرجاع المعلومات التقليدية والآلية على حد سواء، وعلى أساسها تتم المفاضلة بين مختلف أنواع النظم. كما أن قدرة نظم استرجاع المعلومات على إمداد المستفيد بالمعلومات المناسبة من بين الكم الهائل من المعلومات المخزنة في وسائطها الخزنية غاية صعبة المنال، لذا أصبح من ضروريات عمليات التقييم إشراك المستفيد صاحب الحاجة الفعلية للمعلومات في الحكم على صلاحية المعلومات المسترجعة من النظام، فالمعلومات لا قيمة لها إذا لم ترتبط باهتمام أحد المستفيدين أو حاجته الموضوعية .[٦٨]

ويرى الباحث أن تحقيق هدف رضا المستفيد باعتباره أحد معايير تقييم نظام استرجاع المعلومات يبرز دور ترشيح المعلومات كمفهوم، باعتبار ان ترشيح المعلومات تبنى فكرته الأساسية على تقديم مؤشرات مسبقة لنظام المعلومات تمثل الاحتياجات المعلوماتية الفعلية للمستفيد مما يسهم بدرجة كبيرة في تحقيق مخرجات نظام الاسترجاع للحاجات المعلوماتية للمستفيد من جانب، وتقليل زمن تنقية البحث من جانب المستفيد مما ينعكس على مدى رضاه عن النتائج المسترجعة وبالتالي رضاه عن نظام الاسترجاع.

وهناك ثلاث مستويات محتملة يمكن من خلالها وعلى أساسها تقييم نظم استرجاع المعلومات وهي:

- المستوى الأول: تقييم فاعلية النظام التي تتصل باعتبارات رضا المستفيد.
- المستوى الثاني: تقييم فاعلية التكلفة التي تتصل برضا المستفيد عن الكفاءة الداخلية للنظام.

[68] طلال ناظم الزهيري. العوامل المؤثرة في عمليات استرجاع المعلومات.- ص ١.- تاريخ الزيارة ٢٥ ديسمبر ٢٠١١.- متاح في: http://i3.makcdn.com/wp-content/blogs.dir//64868/files//2011/07/d8a7d8b3d8aad8b1d8acd8a7d8b9-d8a7d984d985d8b9d984d988d985d8a7d8aa.doc

- المستوى الثالث: تقييم عائد التكلفة المتصلة بأهمية النظام مقابل تكاليف تشغيله.

وفي سبيل الوصول إلى تقييم موضوعي ودقيق لكفاءة نظم استرجاع المعلومات، لابد من أن يكون للمستفيد دور مباشر فيه من خلال حكمه على نوعية المخرجات التي يحصل عليها بوصفها نتائج للعملية البحثية التي يقوم بها. حيث ما تزال مقاييس الاستدعاء Recall والدقة Precision تحظى بقبول واسع في مجال التقييم النوعي لنظم استرجاع المعلومات، وذلك بسبب ارتباطيهما برضا المستفيد وطريقة حكمه على مخرجات العملية البحثية أولا، و ترابطهما مع بعض ثانياً، حيث يعد معدل الاستدعاء مقياس لقدرة النظام على استرجاع التسجيلات المتصلة بموضوع البحث، بينما يعد معدل الدقة مقياس لقدرة النظام على استبعاد التسجيلات التي ليست ذات صلة بموضوع البحث. [٦٩]

ويرى الباحث أن مستويين من مستويات تقييم نظم استرجاع المعلومات ترتبط ارتباط مباشر بالمستفيد ومدى رضاه عن نتائج الاسترجاع من النظام، لذا فإن تطبيق مفهوم الترشيح في نظم استرجاع المعلومات بوجه عام ونظم استرجاع المعلومات المفتوحة في بيئة الويب بوجه خاص سيقودنا بالتاكيد لتحقيق رضا المستفيد وبالتالي قياس كفاءة نظام الاسترجاع.

حيث أن مفهوم الكفاءة لا يرتبط بقدرة النظام على الاسترجاع فحسب، وإنما في قدرته على إشباع حاجات المستفيدين من المعلومات ذات الصلة بموضوعاتهم البحثية أو ما يعرف اصطلاحا "اكتمال البحث"، فليس من أهداف عمليات تقييم نظم استرجاع المعلومات قياس قدرتها على الاسترجاع باعتبارها عملية نهائية، لان قدرة النظام على الاسترجاع تعد وسيلة المستفيد لاختيار التسجيلات التي يراها وثيقة الصلة بالموضوع. [٧٠]

---

[69] المصدر السابق.

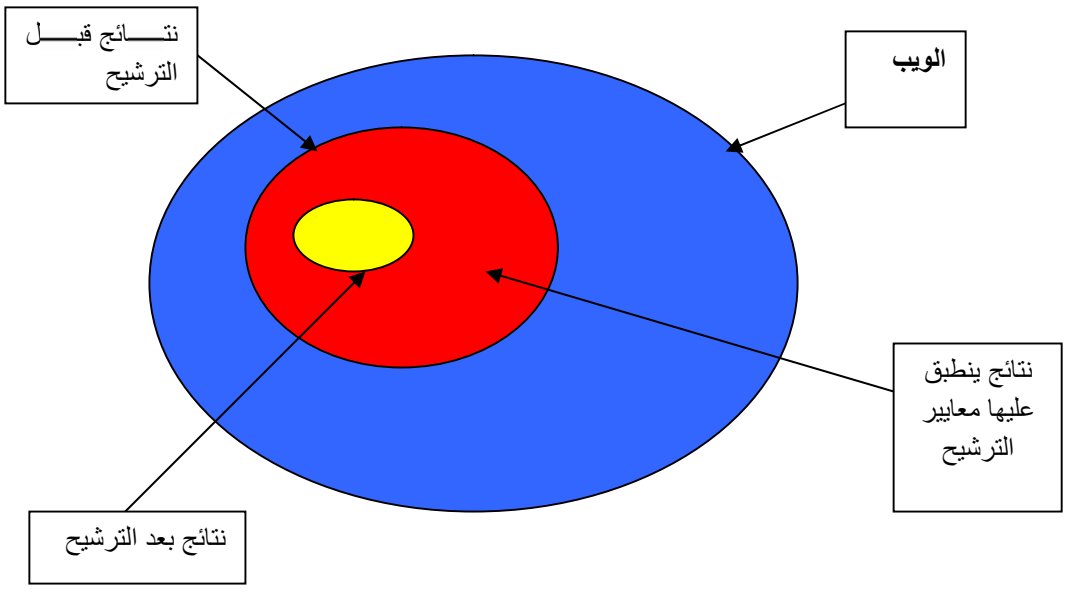
[70] المصدر السابق.

ويرى الباحث من العرض السابق ان مفهوم ترشيح المعلومات المبني على احتياجات المستفيد المعلوماتية يرتبط بقياس مدى كفاءة نظام استرجاع المعلومات، مما يبرز العلاقة بين مفهوم الترشيح ونظم استرجاع المعلومات على مستوى التطبيق والتقييم للنظم.

٦/٢ علاقة الترشيح بنظريات نظم استرجاع المعلومات

سيتم في هذا الجزء من الدراسة مناقشة فرضية أنه يمكن قياس معامل ترشيح المعلومات بنظم الاسترجاع المفتوحة في بيئة الويب استناداً على نظريات استرجاع المعلومات المعروفة في المجال، وذلك من خلال تحليل نظريات استرجاع المعلومات في ضوء متغير الترشيح، وذلك في محاولة للوصول لتصور لقياس معاملات الاسترجاع في البيئة الرقمية في ضوء تطبيق متغير ترشيح المعلومات بنظم استرجاع المعلومات المفتوحة المتمثلة في محركات البحث، وذلك بهدف رسم تصور مقترح لما تمثله النظرية الجديدة لاسترجاع المعلومات المتضمنة لمعيار الترشيح بين معايير الحكم على نظم استرجاع المعلومات المفتوحة.

أستقر الباحث أن التعريف الإجرائي لمعامل الترشيح "نسبة التأثير على نتائج الاسترجاع قبل تطبيق معايير الترشيح التي يحددها المستفيد وبعدها".



شكل رقم (١-١) الترشيح في بيئة الويب

كما يوضح الشكل رقم (١-١)، فأن عملية البحث التي يتم تنفيذها باستخدام محرك البحث وفق معايير ترشيح محددة مسبقاً تتطلب نظرياً مرحلتين في البحث، أولهما الوصول للنتائج قبل الترشيح، وثانيهما تطبيق معايير الترشيح لاسترجاع النتائج التي ينطبق عليها معايير الترشيح، ويمكن توضيح النتائج من خلال عملية بحثية بسيطة في محرك البحث Google باستخدام الكلمات المفتاحية "الربيع العربي"، ويمكن حساب النتائج كالتالي:

معامل الترشيح = النتائج بعد الترشيح / النتائج قبل الترشيح

"الربيع العربي"= 13,000,000 / 27,300,000 = 2.1 تقريبا

نسبة الترشيح = ١٠٠% - (النتائج بعد الترشيح * ١٠٠%) / النتائج قبل الترشيح

قيمة المواد المرشحة (المستبعدة) = قبل الترشيح – بعد الترشيح

قيمة المواد المرشحة (المستبعدة)= 27,300,000 – 13,000,000= 14,300,000

وفق العرض السابق سيسعى الباحث لقياس معامل الترشيح رياضيا من خلال معاملات الاسترجاع في البيئة الرقمية استناداً على نظريتي لانكستر، وكنت.

١/٦/٢ معامل الترشيح وفق نظرية لانكستر

يميل المستفيدون من نظم استرجاع المعلومات للحكم عليها بمعايير التكلفة والوقت ومدى الجودة. ومعايير التكلفة والوقت المتصلة بتقييم نظم استرجاع المعلومات مباشرة على حد بعيد، كما أنها ثابتة نسبياً. أما معايير الجودة فربما كانت

اقل تحديدا أو اقل وضوحا في معالمها، كما أنها تختلف وبشكل ملحوظ تبعا للنظام الذي يتم تقييمه وتبعا لاحتياجات المستفيدين.[٧١]

ويوضح الجدول رقم (١- ١) معاملات الكفاءة عند لانكستر

جدول رقم (١-١) معاملات قياس الكفاءة نموذج لانكستر الملائم لنظم الاسترجاع الثابتة

	صالح للموضوع أ + ج	غير صالح للموضوع ب + د
مسترجع أ + ب	أ استدعاء تحقيق	ب شوشرة
غير مسترجع ج + د	ج فاقد	د استبعاد

اعتمادا على الجدول (١- ١) تكون معاملات قياس الكفاءة كما يلي:

معامل الاستدعاء = الوحدات المسترجعة الصالحة \ كل الصالح للموضوع= **أ \ أ+ج**

معامل التحقيق =الوحدات المسترجعة الصالحة \ كل المسترجع = **أ \ أ + ب**

معامل الشوشرة = الوحدات المسترجعة غير الصالحة \ كل المسترجع = **ب \ أ + ب**

معمل الفاقد = الوحدات غير المسترجعة الصالحة \ كل الصالح = **ج \ أ+ج**

[71] لانكستر، ولفرد و وورنر، آمي. ج. أساسيات استرجاع المعلومات [نظم استرجاع المعلومات] / ترجمة حشمت قاسم.- ط ٣ ، مزيدة ومنقحة . - الرياض : مكتبة الملك فهد الوطنية ، ١٩٩٧ . – ص ص ٢٢٥ – ٢٢٧.

معامل الاستبعاد= الوحدات غير المسترجعة غير الصالحة \ كل غير المتصل= د \ ج+د

ويرتبط الاستدعاء بقدرة النظام على استرجاع الوثائق المتصلة بالموضوع، بينما يرتبط التحقيق بالقدرة على عدم استرجاع الوثائق غير المتصلة بالموضوع، وتعبر نسب الاستدعاء والتحقيق عن قدرة النظام على الترشيح، أي قدرة النظام على استرجاع المطلوب وحجب ما سواه.[٧٢]

ويرى سيد ربيع أن لانكستر بدأ بالحديث عن مقياسي الزمن والتكلفة وهما يعتبران وجها عملة استخدام الويب؛ حيث يترجم الوقت المبذول في البحث إلى تكلفة من جانب الباحث أو تكلفة في الوقت، ولا تختلف هذه المقاييس بين جميع نظم وخدمات استرجاع المعلومات. إلا أن تقرير الكفاءة يتوقف على الوحدات المسترجعة من جانب النظم لإفادة الباحثين؛ فإذا ما كانت هذه الوحدات تلبي الجانب الأكبر من احتياجات المستفيد أصبحت التكلفة في مستوى منخفض إلى جانب الوقت المبذول لاسترجاع تلك الوحدات، وتسمى الوحدات المسترجعة المطابقة لاحتياجات المستفيد بالتحقيق؛ أحد أهم عوامل الحكم على كفاءة الاسترجاع. وبالتعرض إلى تفسير نظام استرجاع لانكستر للمعلومات يتضح أن مفهوم الاستدعاء الذي يعتمد عليه المستفيد في تحديد كفاءة النظام يختلف عن ما يوضحه الجدول (١- ١) لنظام معاملات الكفاءة للانكستر؛ حيث ينظر المستفيد عند بحث نظم الاسترجاع التقليدية وعلى الويب، إلى أن الاستدعاء هو كل ما تم استرجاعه من داخل قاعدة البيانات ككل: أي كل المسترجع. على حين أن معامل الاستدعاء عند لانكستر يتم حسابه اعتمادا على معرفة عدد الوحدات المسترجعة ذات الصلة بالموضوع (وهو ما لا يمكن إدراكه داخل محركات البحث سريعة التحديث)، وعلى ذلك فإن الاستدعاء عند لانكستر هو فقط مجموعة من كل المسترجع إلا أنه يساوي كل المسترجع عمليا عند المستفيد.

٢/٦/٢ نماذج قياس كفاءة نظم الاسترجاع على الويب

[72] المصدر السابق. ص ص ٢٣٠– ٢٣١.

وقد قدم الباحث سيد ربيع نموذج لقياس كفاءة نظم الاسترجاع على الويب تعتمد على ما يلي:

إن اعتماد نظم استرجاع المعلومات التقليدية على نظام لانكستر يختلف بدوره عما يمكن التعامل معه في نظم استرجاع الويب، وذلك إنما يرجع إلى الديناميكية التي تعمل بها نظم استرجاع الويب في إضافة آلاف المواقع يوميا إلى قواعدها، فضلا عن اعتماد أساليب مختلفة في التنظيم والبحث لصفحات الويب وملفات الصور، إلى جانب البعد التام عن التدخل البشري في تنظيم وبحث محركات بحث الويب. ويجب ملاحظة وجود الاستدعاء في النوع السابق كجانب مكمل للفاقد لتكملة كل المتصل بالموضوع الذي لم يسترجع، في حين أن الناحية العملية تحتم إطلاق مصطلح الاستدعاء على كل الوحدات المسترجعة لكي تلاءم التطبيق العملي لحكم المستفيدين. ويطرح الباحث النموذج التالي الممثل في الجدول رقم (١-٢) للاستخدام في تحديد كفاءة نظم استرجاع الويب عامة، كما يلي :-

جدول رقم (١-٢) معاملات قياس الكفاءة المقترحة لنظم استرجاع الويب

	صالح للموضوع أ + ج	غير صالح للموضوع ب + د
مسترجع (الاستدعاء) أ + ب	أ تحقيق	ب شوشرة
غير مسترجع د	د استبعاد	

وتكون المعاملات كما يلي :-

معامل الاستدعاء = كل المسترجع \ كل ما يشمله النظام = أ + **ب** \ أ + **ب** + د

معامل التحقيق = كل المسترجع الصالح \ كل الاستدعاء (المسترجع) أ \ أ + **ب**

معامل الشوشرة= كل المسترجع غير الصالح\ كل الاستدعاء (المسترجع) **ب** \ أ + **ب**

معامل الاستبعاد =كل المستبعد \ كل ما يشمله النظام = د \ أ + ب + د

تم بناء النظام السابق على أسس هي:-

١. إن من يصنع التنظيم داخل محركات البحث هو برنامج (نظام المعلومات) ومن غير المؤكد معرفة عدد الوحدات التي تلبي موضوع المستفيد.

٢. إن استخدام مصطلح الفقد مع الاسترجاع في النظم الآلية لا يوافق بيئة الويب التي تعمل على تكشيف كل كلمات النص التي تصبح نقاط استرجاع للنص، ومن ثم فإن التحكم في استدعاء النصوص يرجع إلى المستفيد المحدد للموضوع. واعتمادا على عدم تحديد النظام لما يحويه من وحدات عن موضوع بعينه فإن حساب الفقد أو الفاقد من الوحدات ليس له معنى.

٣. إن طبيعة النظم الآلية تبنى على استدعاء كل ما يطابق إستراتيجية البحث ومن ثم لا يعطي النظام من جانبه أي شوشرة، وإنما يحدد المستفيد تلك الشوشرة من وجهة نظر احتياجات خاصة له. حيث يسترجع النظام كل ما تستدعيه كلمات البحث من وحدات.

٤. إن حقيقة التعامل مع نظم المعلومات الآلية هي أن الاستدعاء هو كل المسترجع وأن التحقيق هو ما يطابق احتياج المستفيد من داخل الاستدعاء.

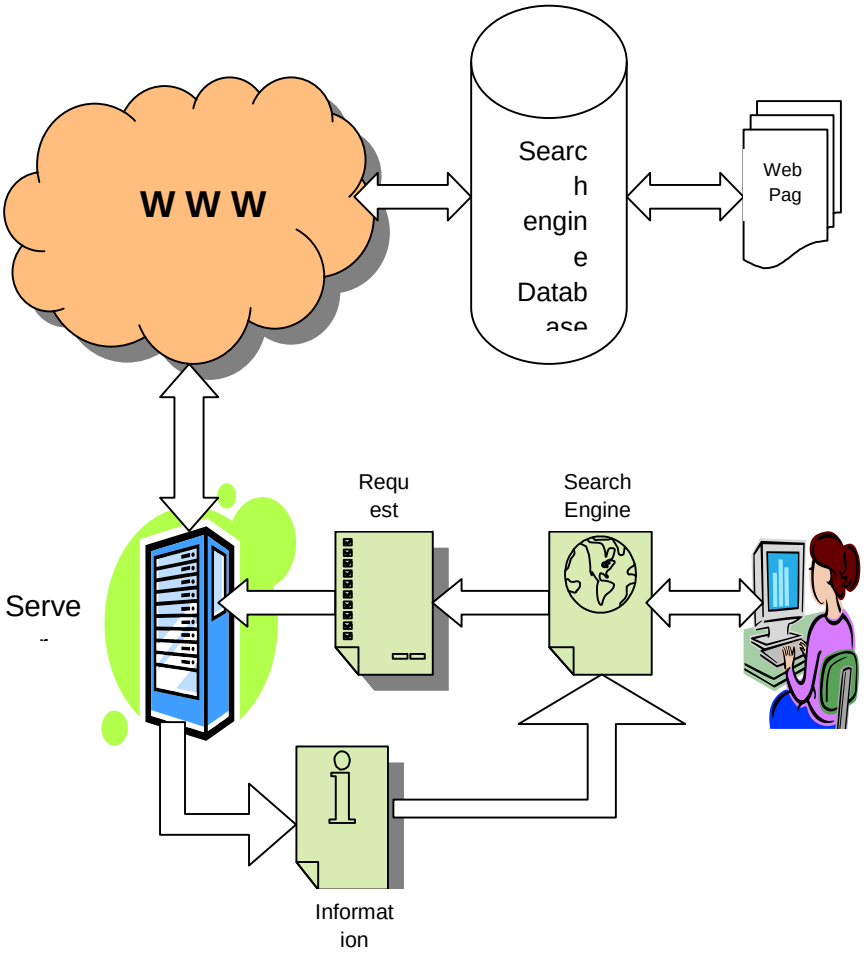
ويتعامل النموذج السابق مع طبيعة نظم استرجاع المعلومات على الويب التي تتعامل مع الموضوعات في صفحات الويب كما قام عليه مصممو هذه الصفحات. حيث ينتج عن تكشيف محركات البحث للنص الكامل أن تكون مختلف كلمات النص مداخل وكلمات مفتاحيه يتم الاسترجاع من خلالها.[٧٣]

٣/٦/٢ معامل الترشيح وفق نظرية لانكستر

ومن العرض السابق لنظريات نظم الاسترجاع في النظم الثابتة والمفتوحة (الويب) يرى الباحث أن الترشيح كعملية تعد متغير من المتغيرات التي تؤثر في عملية الاسترجاع على الويب من خلال وضع شروط ومعايير يتم في ضوءها تنفيذ

[73] سيد ربيع سيد ابراهيم. مصدرسابق. ص ص ٢١٩ – ٢٢١.

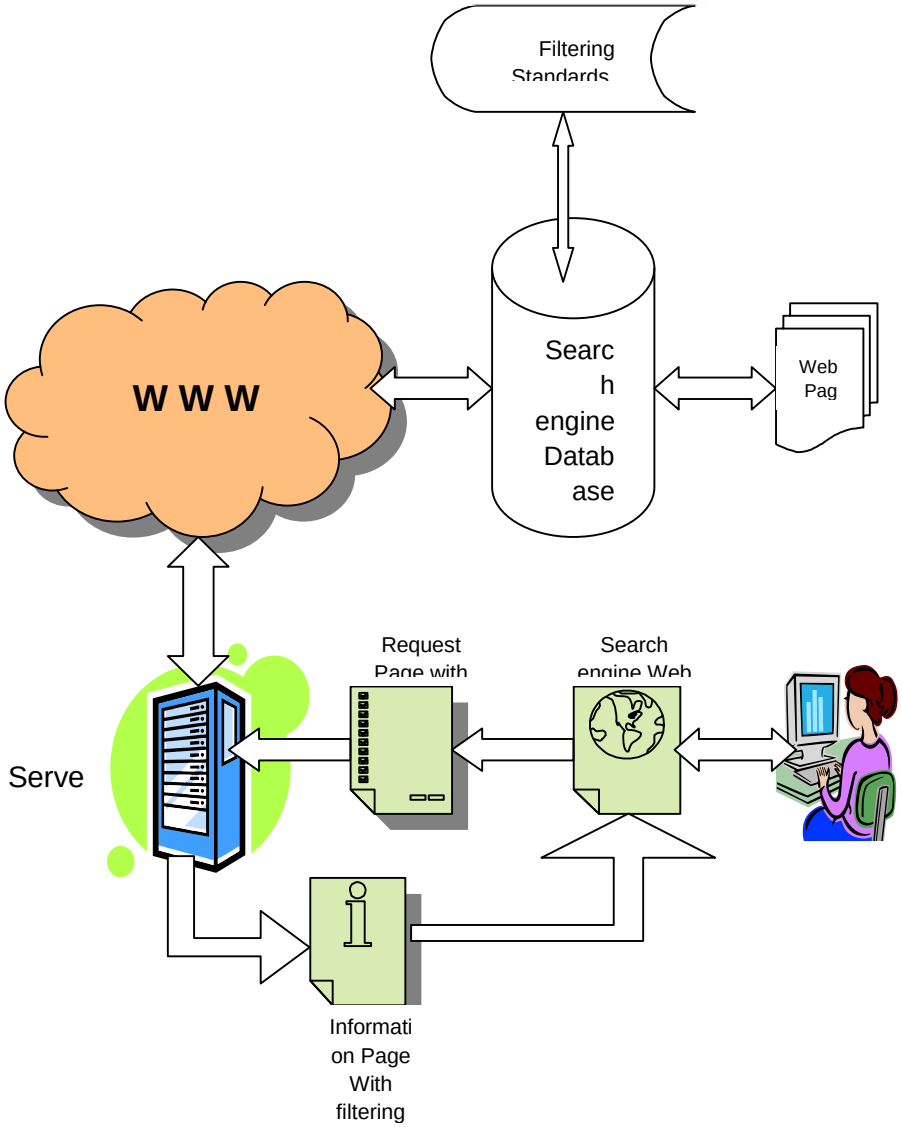
عملية الاسترجاع. لذا سيوضح الباحث تصوره لعملية الاسترجاع في بيئة الويب في حالتي تطبيق وعدم تطبيق معايير الترشيح.



شكل رقم (١-٢) الاسترجاع من الويب باستخدام محركات البحث بدون ترشيح

ويوضح الشكل (١-٢) مراحل تقديم الاستفسار البحثي لنظام استرجاع مفتوح يمثله محرك البحث دون تنشيط خاصية الترشيح المتوفرة بمحرك البحث، ومن

الملاحظ أن النتائج يتم استرجعها مباشرة من قاعدة البيانات لتوجه للمستفيد ليقرر مدى ملاءمتها وفق معاييره واحتياجاته المعلوماتية.



شكل رقم (١-٣) الاسترجاع من الويب باستخدام محركات البحث مع تنشيط خاصية الترشيح

في حين يوضح الشكل (١-٣) مراحل تقديم الاستفسار البحثي لنظام استرجاع مفتوح يمثله محرك البحث مع تنشيط خاصية الترشيح المتوفرة بمحرك البحث والمبنية

على معايير خاصة بمحرك البحث ولا يتدخل فيها المستفيد، ومن الملاحظ أن النتائج لا يتم استرجعها مباشرة من قاعدة البيانات لتوجه للمستفيد، حيث تمر نتائج البحث في قاعدة البيانات على مرحلة وسيطة يتم فيها تطبيق معايير الترشيح الموضوعة مسبقاً ليتم استبعاد النتائج المسترجعة وتنطبق عليها معايير الترشيح الخاصة بمحرك البحث، ثم يتم ارسال النتائج التي لاينطبق عليها معايير الترشيح للمستفيد ليقرر مدى ملاءمتها وفق معاييره واحتياجاته المعلوماتية.

ويرى الباحث أن الترشيح كعملية في محرك البحث الهدف منها استرجاع نتائج من الويب تضم مواد لا تنطبق عليها معايير معينة موضوعة مسبقا بمحرك البحث؛ وهذه المواد المسترجعة قد تناسب الاحتياجات البحثية للمستخدم وقد لا تناسبها، وتؤدي هذه العملية في الوقت ذاته إلى استبعاد مواد تنطبق عليها هذه المعايير الموضوعة مسبقاً؛ وهذه المواد المستبعدة قد تناسب الاحتياجات البحثية للمستخدم وقد لا تناسبها، لذا فأن معامل الترشيح يعد متغير يؤثر في عملية الاسترجاع من بيئة الويب، لذلك فهناك حالتان للاسترجاع من خلال محرك البحث في بيئة الويب؛ الأولى في حالة الاسترجاع بدون تنشيط خاصية الترشيح والثانية الاسترجاع مع تنشيط خاصية الترشيح. لذا من الضروري إلقاء الضوء على عدد من المفاهيم المرتبطة بهذا الموضوع.

يرى الباحث أنه لكي يتم وضع تصور لنظرية استرجاع المعلومات من الويب بواسطة محرك البحث في ضوء متغير الترشيح يجب إيضاح ما يلي:

معايير الترشيح: هي معايير موضوعة مسبقا بمحرك البحث لاستبعاد المواد التي تعد مواد غير ملائمة من نتائج البحث، وهي معايير غير معلنة بوضوح في محرك البحث.

المواد غير الملائمة: هي المواد التي تنطبق عليها معايير الترشيح بمحرك البحث، وهي تمثل وجهة نظر واحدة وهي وجهة نظر القائمين على محرك البحث وليس بالضرورة أن تتطابق مع رأي المستفيد.

لذا سيتم حساب معامل الترشيح وفق آلية تطبيقه في محرك البحث في ضوء المقارنة بين حالتي الاسترجاع السابقتين باعتبار أن المتغير الأساسي في عملية الاسترجاع الحالية هي تنشيط أو عدم تنشيط خاصية الترشيح أو بمعنى أخر تطبيق معايير الترشيح أو عدم تطبيقها.

وسيتم استبعاد رأي المستفيد فيما يتعلق بمعاملات نظرية الاسترجاع نظرا لأن قضية الترشيح وما يجب ترشيحه هي قضية نسبية تختلف من مستخدم لأخر وفق معايير ترشيحه الخاصة التي ترتبط باحتياجاته المعلوماتية واهتماماته، كما أنها أحيانا تكون نسبية للمستخدم الواحد من وقت لأخر بتغير اهتماماته الموضوعية، لذا فقد وضعت خاصية الترشيح اختيارية بمحرك البحث بحيث يمكن للمستخدم تنشيطها أو عدم تنشيطها حسبما يرى، وفي ضوء ذلك يمكن عرض المعاملات التالية:

المسترجع = يمثل كل النتائج المسترجعة من الويب

غير مسترجع = يمثل النتائج غير المسترجعة من الويب

بدون ترشيح = هي المواقع المسترجعة ولا ينطبق عليها معايير الترشيح

الترشيح= هي المواقع المسترجعة وتنطبق عليها معايير الترشيح

تحقيق= النتائج المسترجعة ولم تنطبق عليها معايير الترشيح (النتائج بعد الترشيح)

شوشرة= النتائج المسترجعة و تنطبق عليها معايير الترشيح (جزء من النتائج قبل الترشيح)

استبعاد= النتائج غير المسترجعة سواء تنطبق أو لا تنطبق عليه معايير الترشيح

فاقد مرشح = النتائج غير المسترجعة ولا ينطبق عليها معايير الترشيح

جدول رقم (١-٣) معاملات قياس الترشيح المقترحة بنظم استرجاع الويب

	بدون ترشيح	ترشيح	المجموع

مسترجع	أ شوشرة	ب تحقيق	أ + ب
غير مسترجع	ج استبعاد	م فاقد مرشح	ج + م
المجموع	أ + ج	ب + م	

ووفق جدول (١-٣) يمكن التعبير عن معاملات الاسترجاع بالمعادلات التالية:

تحقيق = مسترجع مرشح / كل المرشح ب = ب / ب +م

شوشرة = مسترجع غير مرشح / كل المسترجع أ = أ / أ + ب

استبعاد = غير مسترجع غير مرشح / كل غير المرشح ج = ج / أ + ج

فاقد مرشح = غير مسترجع مرشح / كل المرشح م = م / ب + م

وتكمن المشكلة الأساسية في تطبيق هذه المعاملات رياضيا وفق نظرية لانكستر وجود العديد من القيم المجهولة في المعادلة، والمتمثلة في كل النتائج غير المسترجعة والتي يمثلها الاستبعاد، والتحقيق الذي بني على معايير ترشيح لايتبناها المستفيد، وذلك في نظم استرجاع المعلومات المفتوحة التي يصعب السيطرة على مدخلاتها وبالتالي يصعب حساب الفاقد والمستبعد، كما أن القياس على نظم الاسترجاع المفتوحة التي تعتمد معايير ترشيح خاصة بها دون النظر لاحتياجات المستفيدين المعلوماتية واشتراطاتهم في الترشيح سيعطي مؤشرات خاطئة غير معبرة عن حقيقة معامل الترشيح.

٤/٦/٢ معامل الترشيح وفق نظرية كنت

وضع كنت وزملاؤه قانونا تعتمد فكرته عل أن أي نظام لاسترجاع المعلومات يحتوي على عدد من الوثائق تمثل في مجموعها "ن" وأن هناك عدد "م" وثيقة يحتمل أنها تقع في مجال اهتمام المستفيد. وأن هناك عدد أقل من الوثائق "ف" يمثل الاحتياج الفعلي للمستفيد.

وعلى ذلك تكون:

ن = عدد جميع الوثائق التي يضمها النظام.

م = عدد الوثائق التي يحتمل أنها تقع في مجال اهتمام المستفيد.

ف = عدد الوثائق التي تمثل الاحتياجات الفعلية للمستفيد.

وقد عرف كنت وزملاؤه مجموعة من المعاملات تعرف نظام البحث في قاعدة بيانات، وبها يمكن قياس البحث والاسترجاع وذلك كما يلي:

معامل الكثافة = نسبة الوثائق التي تقع في مجال اهتمام المستفيد = عدد الوثائق التي يحتمل أنها تقع في مجال اهتمام المستفيد/ عدد الوثائق الكلي = م / ن

معامل الاستبعاد = نسبة الوثائق المستبعدة = عدد الوثائق الكلي – عدد الوثائق التي يحتمل أنها تقع في مجال اهتمام المستفيد / عدد الوثائق الكلي = ن – م / ن

معامل الاستبقاء = نسبة الوثائق التي تمثل احتياجات المستفيد الفعلية = عدد الوثائق التي تمثل الاحتياجات الفعلية للمستفيد / عدد الوثائق التي يحتمل أنها تقع في مجال اهتمام المستفيد = ف / م

معامل التشويش = نسبة الوثائق التي تم استرجاعها ولا تقع ضمن احتياجات المستفيد = عدد الوثائق التي يحتمل أنها تقع في مجال الاهتمام – عدد الوثائق التي تمثل الاحتياجات الفعلية / عدد الوثائق التي يحتمل أنها تقع في مجال اهتمام المستفيد = م – ف / م

وحيث لا يوجد نظام استرجاع نموذجي فإنه يظل هناك احتمال دائما أن هناك عدد من الوثائق تقع في مجال اهتمام المستفيد ولم يتم استرجاعها وهي "س"

معامل الفاقد = نسبة الوثائق التي تقع في مجال اهتمام المستفيد ولم يتم استرجاعها = عدد الوثائق التي تقع في مجال اهتمام المستفيد ولم يتم استرجاعها / عدد الوثائق الكلي = س / ن.[٧٤]

ويرى الباحث أن نظرية كنت وزملاؤه تصلح للتطبيق فقط على نظم الاسترجاع المغلقة مثل قواعد البيانات أما النظم المفتوحة مثل بيئة الويب وتحديدا قواعد بيانات محركات البحث والتي يتم تحديثها بالإضافة أو الحذف منها بشكل مستمر فيصعب قياس هذه المتغيرات فيها، هذا إلى جانب متغير أخر يؤثر في عملية الاسترجاع ألا وهو تطبيق متغير الترشيح في بيئة الاسترجاع، لذا يقترح الباحث رؤية لنظرية كنت وزملاؤه في بيئة الويب وفق متغير الترشيح وهي كما يلي:

ن = النتائج المسترجعة قبل الترشيح (لأنها تمثل جميع الوثائق التي يضمها النظام)

م = النتائج بعد الترشيح التي لم يطبق عليها معايير الترشيح بالنسبة لمحرك البحث

ف = النتائج بعد الترشيح التي انطبقت عليها فعليا معايير المستفيد للترشيح

لا يمكن "**ف**" حسابها عمليا نظرا لاشتمالها على عدد من المتغيرات المرتبطة بشبكة الويب والمستفيد والتي يصعب معرفتها، ومنها:

- تختلف معايير الترشيح للمستخدم عن معايير الترشيح لمحرك البحث
- تختلف معايير الترشيح من مستخدم لأخر نظرا لتنوع وجهات نظر المستفيدين فيما يجب ترشيحه وما لا يجب ترشيحه
- تختلف معايير الترشيح للمستخدم الواحد من وقت لأخر وفق احتياجاته.

[74] طلال ناظم الزهيري. مصدرسابق.

- معامل الكثافة = بعد الترشيح / قبل الترشيح = م / ن
- معامل الاستبعاد (الترشيح) = (قبل الترشيح – بعد الترشيح) / قبل الترشيح = (ن - م) / ن

أما معاملات الاستبقاء والتشويش والفقد فلا يمكن حسابها في بيئة الويب وفق معايير الترشيح

٥/٦/٢ دقة الاسترجاع وترشيح المعلومات

تقاس كفاءة التكشيف (لغة الاسترجاع) في نظم الاسترجاع على أساس معدل الاستدعاء ومعدل الدقة، والمعدل الأمثل لأي منهما هو (١) والذي يعتبر تحقيقه صعباً، ولابد من تحقيق الكفاءة المثلى بأقل التكاليف وأكبر منفعة للمستفيد مع مراعاة ألا تزيد تكلفة المعالجة على الفائدة المتوقعة للمستفيد سواء من حيث الجهد أو المال.

ويقاس نجاح خدمات المعلومات بقياسين نوعيين رئيسيين، هما:

- حصول المستفيد على المعلومات التي يبحث عنها.
- مدى اكتمال ودقة نتائج البحث.[٧٥]

ونظرا للعلاقة بين مفهوم الترشيح ودقة الاسترجاع تناول الباحث نظريات استرجاع المعلومات - وتحديداً نظريتي " لانكستر وكنت "- ومفهوم دقة الاسترجاع من حيث معاملات الاستدعاء والتحقيق للنتائج المسترجعة ومدى رضا المستفيد عن هذه النتائج، وقد نوقشت هذه النظريات في ضوء معايير الدقة والكفاءة لنظم استرجاع المعلومات؛ وتوصل الباحث إلى أنه نظرياً يمكن حساب دقة معامل الترشيح وفق آلية تطبيقه في محرك البحث ودون الأخذ في الاعتبار حاجات المستفيد نظراً لأن معايير الترشيح لدى المستفيدين نسبية وتختلف من مستفيد لأخر وتتغير معايير المستفيد نفسه بمرور الوقت ووفق احتياجاته؛ ولكن عمليا هناك العديد من الحقائق التي يجب توفرها

[75] سناء حافظ التكراوي. نظم استرجاع المعلومات : بين النظرية والتطبيق.- عمان: دار صفاء للنشر،٢٠٠٤.- ص ص ١٢٦- ١٢٧.

حتى يمكن أن يتحقق هذا القياس بدقة ومنها؛ معرفة معايير الترشيح الموضوعة بمحرك البحث وبدقة؛ وهل هذه المعايير ثابتة أم متغيرة؟ وذلك لتحديد مفهوم المواد غير الملائمة.

لذا فإذا تم قياس دقة نتائج الاسترجاع لمحرك البحث في ضوء متغير الترشيح الذي بني على معايير غير واضحة ووفق المعايير السابق ذكرها لقياس مدى نجاح خدمات المعلومات؛ فإن معياري حصول المستفيد على المعلومات التي يبحث عنها ومدى اكتمال ودقة نتائج البحث لا يمكن قياسهما؛ نظراً لأن الترشيح قد يحجب معلومات يرغب المستفيد في الحصول عليها وبالتالي لن يشعر المستفيد بدقة، واكتمال، والرضا عن النتائج المسترجعة نظراً لمعرفته بحجب معلومات قد تكون ذات صلة بمجال اهتمامه.

لذا يرى الباحث أنه لقياس دقة الاسترجاع في بيئة الويب من خلال محركات البحث وفي ضوء متغير الترشيح هناك اتجاهان؛ الأول انه يجب في المقام الأول تحديد وبدقة معايير الترشيح، حتى يوائم المستفيد احتياجاته وهذه المعايير المؤثر في نظام استرجاع محركات البحث ليتحقق في النهاية هدف عملية الاسترجاع المتمثل في رضا المستفيد عن النتائج وهو ما يمكن تطبيقه نظرياً؛ حيث لا يمكن تطبيق هذه الميزة عملياً على محركات البحث نظراً لتنوع فئات وثقافات واحتياجات مستخدميها فهي بمثابة أدوات بحث عالمية فكيف يمكن وضع معايير عالمية تتوافق وهذا المجتمع المتنوع من مستخدمي محركات البحث، أما الاتجاه الثاني أن تكون معايير الترشيح متغيرة ويكون للمستفيد القدرة على التعديل والتغيير سواء بالحذف والإضافة إلى هذه المعايير وبالتالي يستطيع المستفيد أن يحدد احتياجاته مما سينعكس على رضاه عن النتائج المسترجعة؛ وهذه الميزة متوفرة فعلياً في عدد من برامج المرشحات المتوفرة وهو ما يمكن تطبيقه نظرياً على محركات البحث ولكن عملياً يتطلب الكثير من الجهد في الإعداد والتنظيم. وفي النهاية هناك عامل أخر أكثر أهمية سيؤثر على كلا الاتجاهين السابقين وهو أن الجهات المسئولة عن محركات البحث هي مؤسسات ربحية وليست خدمية وبالتالي لا يوجد مبرر أو عائد مادي يعود عليها من وراء ذلك.

١/٣ نتائج الدراسة

انطلقت الدراسة من فرضيتين، أولهما أن ترشيح المعلومات يمثل شكلا جديدا من أشكال أسترجاع المعلومات تطبق آلياته في نظم استرجاع المعلومات المفتوحة، لإثبات هذه الفرضية استعرض الباحث ادبيات الانتاج الفكري التي تعالج مفهومي الترشيح واسترجاع المعلومات بهدف إبراز العلاقة بينهما؛ وقد توصل الباحث إلى أن ترشيح المعلومات هو الآلية الأكثر دقة لاسترجاع المعلومات في بيئة الويب نظراً لاهتمام الترشيح باهتمامات المستفيد الممتدة وليست الآنية.

أما الفرضية الثانية المتعلقة بإمكانية قياس معامل ترشيح المعلومات بنظم الاسترجاع المفتوحة في بيئة الويب استناداً على نظريات استرجاع المعلومات المعروفة في المجال، ولاثبات هذه الفرضية تعرض الباحث بالتحليل لنظريتي لانكستر، وكنت في ضوء متغيرات بيئة الويب التي فرضتها على طبيعة نظم الاسترجاع، كما ناقش الباحث آلية قياس معاملات الاسترجاع في بيئة الويب في محاولة لقياس معامل الترشيح، والذي ثبت إمكانية قياسه نظرياً في حال توافق معايير الترشيح مع احتياجات المستفيد الفعلية وليس وفق التطبيق الحالي للترشيح بمحركات البحث.