مجلة بحوث
كلية الآداب

البحث (٢٦)

دراسة مقارنة بين محرك بحث GOOGLE ونظم الأرشفة الدلالى VIVO

إعداد

الباحثة / مرثا نبيل صبرى

لدرجة الماجستير بقسم المكتبات والمعلومات

كلية الآداب - جامعة المنوفية

السنة ٢٧ العدد (١٠٧) أكتوبر ٢٠١٦م

http : // Art.menofia . edu. eg *** E- mail: rifa2012@ Gmail.com

الممسوحة ضوئيا بـ CamScanner

دراسة مقارنة بين محرك بحث GOOGLE ونظام الأرشفة الدلالي VIVO [1]

الباحثة/ مـرثا نبيل صبري.

لدرجة الماجستير بقسم المكتبات والمعلومات كلية الاداب – جامعة المنوفية

المستخلص

هدفت الدراسة إلي التعرف علي الوضع الراهن لنظم الأرشفة الرقمية الدلالية (SSDA) **Semantic Digital Archives System** المتاحة علي الشبكة العنبكوتية من خلال تحليل أهم نظم الأرشفة الرقمية الدلالية المتاحة **VIVO**[2]، ودراسة بنيتها التكوينية وخوازميات عملها بهدف التعرف علي أبرز سماتها و ما يميزها عن محرك البحث **GOOGLE**[3].

إعتمدت الدراسة على المنهج الوصفي أسلوب البحث المقارن لمقارنة أداء نظام الأرشفة الطبي الدلالي **VIVO** بمحرك البحث جوجل **GOOGLE** من حيث الإعتمادية والوظيفة ،ومن أهم النتائج التي توصلت إليها الدراسة أن نظم الأرشفة الدلالية **Semantic Digital Archive System** (SDAS تختلف عن نظيرتها التقليدية في عدة جوانب من بينها تقنيات تطوير النظم ، وآلية عملها ، والنتائج التي تقدمها للمستفيد حيث توفر نظم الأرشفة الرقمية الدلالية إمكانية لإكتشاف المعرفة ،والتنقيب عن البيانات، وآليات ربط وبحث المفاهيم المرتبطة،كما أوصت الدراسة إجراء المزيد حول تطبيقات الويب الدلالي في مجال الوثائق والأرشيف.

[1] هذا البحث نتائج بحث رسالة ماجستير مستله بعنوان :
مرثا نبيل. نظم الأرشفة الرقمية الدلالية:دراسة تحليلية.-رسالة (ماجستير)، جامعة المنوفية،٢٠١٦.

[2] http://www.vivoweb.org/

[3] https://www.google.com.eg/

الممسوحة ضوئيا بـ CamScanner

0/1 المقدمة:

تعتبر شبكة الويب أغنى المصادر المعلوماتية بما تحويه من معلومات ومصادر متنوعة يمكن الوصول إليها عن طريق محركات البحث التقليدية، غير أن تنظيم هذه المعلومات والمستندات بصورة تُسهل عملية البحث فيها والوصول إليها، يعتبر أمراً غاية في الصعوبة في ظل التزايد المستمر في حجم المعلومات المنشورة في شبكة الويب، حيث مرت الشبكة العنكبوتية بأجيال عده منذ أن تم تطورها علي يد تيم بيرنزر لي **Tim Berners- Lee** ، فقد دعمت في النشر الألكتروني في من خلال صفحات تستخدم لغة ترميز النصوص الفائقة **HTML** ،ثم تطورت مع ظهور الجيل الثاني من الويب تم إستخدام تقنيات وبروتوكولات حديثة للنشر الألكتروني، وبدأت عملية إستخدام الشبكة العنكبوتية تتجه نحو التفاعل بين المستفيد والنظام من خلال التقنيات المختلفة مثل الويكي **Wiki** وخدمة خلاصات المواقع **RSS**، والشبكات الاجتماعية ...إلخ ومع الأتجاه نحو الجيل الثالث للويب ظهر ما يعرف ظهر ما يعرف بتقنيات الويب الدلالي **Semantic Search**، ونظم الأرشفة الرقمية التي تقدم هذا النوع من البحث.

1/1 ظاهرة الدراسة:

تتمركز مشكلة الدراسة حول:

التحقق من مدى إمكانية وفاعلية معايير الويب الدلالي المتمثلة في لغة أنطولوجيا الويب **Ontology Web Languague (OWL)** ،لغة التكويد قابلة الامتداد **Extensible Markup Language (XML)** ونماذج إطار وصف المصادر **Resource Description Framework (RDF)** ، مع نظام الأرشفة الدلالي **VIVO** ومقارنة أدائه في البحث والإسترجاع بأحد أشهر محركات البحث وهو **GOOGLE** .

2/1 أهداف الدراسة:

تسعي الدراسة إلي تحقيق الأهداف علي النحو التالي:

١. إلقاء الضوء علي مفهوم وخصائص الويب الدلالي الذي يمثل أحد الإتجاهات الحديثة في مجال إسترجاع المعلومات.

الممسوحة ضوئيا بـ CamScanner

١. التعرف على خصائص أكثر التطبيقات تطورا في مجال الأرشيف الرقمي من حيث البنية والتطبيقات ودراسة مدى فاعليتها في تنظيم التخزين والاسترجاع والبحث.

٢. تعرف أبرز متطلبات الويب الدلالي وتقنياته.

٣. مقارنة أداء نظام الأرشفة الدلالي المقترح VIVO بمحرك البحث جوجل GOOGLE من حيث الإعتمادية والوظيفية.

٤. التعرف على الفرق بين نتائج نظام الأرشفة الدلالي المقترح VIVO و نتائج محرك البحث GOOGLE

3.1 تساؤلات الدراسة:

١. ما مفهوم وخصائص الويب الدلالي ؟

٢. ما أبرز متطلبات تطبيق الويب الدلالي ؟

٣. ما الفرق بين نتائج بحث نظام الأرشفة الدلالي المقترح VIVO و نتائج بحث محرك البحث GOOGLE ؟

٤. ما نسبة الفرق بين أداء نظام الأرشفة الدلالي المقترح VIVO و محرك البحث GOOGLE

4.1 أهمية الدراسة :

تم توضيح أهمية الدراسة على النحو الآتي:

١. تكمن أهمية هذه الدراسة في أنها تعد من الدراسات العربية القليلة التي تتجه نحو موضوع البحث الدلالي الذي يمثل ثورة حقيقية في نظم استرجاع المعلومات، ويمكن أن تفيد هذه الدراسة في توجيه الاهتمام العربي بشكل أكبر نحو تطبيقات الجيل الثالث من الويب، واستثمار هذا التطور التقني في مجال البحث واسترجاع المعلومات.

الممسوحة ضوئيا بـ CamScanner

5/1 منهج الدراسة وأدواته:

إعتمدت الدراسة على المنهج الوصفي كإطار عام لرصد وجمع البيانات حيث يساعد في الوصف التام تفصيلاً وتدقيقاً ، واعتمدت الباحثة في إجراء الدراسة علي أسلوب البحث المقارن **Comparative Method** كما تبناها د. **محمود الجندي**[٤] ضمن تصنيفه لمناهج البحث العلمي و أساليبه وذلك لمقارنة أداء نظام الأرشفة الطبي الدلالي **VIVO** بمحرك البحث جوجل **GOOGLE** من حيث الإعتمادية والوظيفة عن طريق دراسة أدبيات الموضوع من بحوث ودراسات، من خلال البحث فى قواعد البيانات، ومواقع الإنترنت ذات العلاقة بموضوع الدراسة.

- **أدوات الدراسة:**

تم إستخدام قائمة مراجعة يجيب عليها أعضاء هيئة التدريس بكلية الطب جامعة المنوفية لجمع البيانات حول موضوع الدراسة، للتعرف ميولهم و إتجاهاتهم نحو إستخدام الويب الدلالية نظام الأرشفة الطبي الدلالي **VIVO** مقارنتهً بمحرك البحث جوجل **GOOGLE**.

عينة الدراسة:

بلغ عدد أعضاء هيئة التدريس بكلية الطب جامعة المنوفية **1223** القائمين علي رأس العمل منهم **363** قائمين بأجازة خاصة ، **860** قائمين بالعمل بالجامعة ،لهذا تناولت عينة الدراسة **10%** من الأطباء القائمين بالعمل بالجامعة تمثلت في **86** عضو للتعرف علي ميولهم وإتجاهاتهم في إسترجاع المعلومات علي نظم الأرشفة الدلالية **VIVO** كنموذج طبي ومقارنتها بمحرك البحث **GOOGLE**، وقد تم إختيار عينة الدراسة من أعضاء هيئة التدريس كليه الطب لأن النظام الدلالي **VIVO** طبي، ويرجع السبب في مقارنه محرك البحث**Google** بنظام **VIVO** الدلالي هو أن **Google** يقدم أكثر بكثير من محرك بحث ويفضل إستخدامه العديد من الباحثين حيث يقوم بعرض نتائج البحث مرتبة حسب

[٤] محمود عبد الكريم الجندي. مناهج البحث في مقالات دوريات المكتبات و المعلومات العربية : دراسة تحليلية. مجلة مكتبة الملك فهد الوطنية. مج.١٨ ، ع.٢ (مايو-نوفمبر ٢٠١٢)، ص. ٣١٥

الممسوحة ضوئيا بـ CamScanner

الأهمية والتاريخ والأثر العلمي الذي تركته في مجال التخصص، وبالتالي يتيح لنا التوصل لأهم ما تم نشره والاطلاع عليه بسهولة ويسر، وإمكانية الوصول إلى النصوص الكاملة للأبحاث من خلال PDF,DOC مما يجعل جوجل متخصص.

6/1 مصطلحات الدراسة:

نظام الأرشيف الرقمي Digital Archive System: عرفه قاموس ODLIS بانه نظام رقمي يسمح بتخزين وتوفير سبل الوصول إلى المواد الرقمية على المدى الطويل لضمان أن المواد لا تزال قابلة للاستخدام في ظل التغيرات التقنية [5]، كما يذكر Dollar تعريفاً آخر بأنها تلك النظم التي تعتمد علي تكنولوجيا الحاسبات،والتي تتيح العمليات والإجراءات ممن أجل الإدارة والسيطرة علي عمليات تخزين الوثائق الإلكترونية وإسترجاعها[6]

محركات البحث Search Engines: عبارة عن أداة تقوم بالبحث عن المعلومات على الإنترنت وتخزين عناوينها على مرصد البيانات الخاص بها، ثم تقوم بإتاحتها للمستفيدين كل حسب المصطلح أو المصطلحات المستخدمة في البحث ومن ثم تمكن المستفيد من الوصول إلى مصادر المعلومات المختلفة على الإنترنت ،ويتم تجميع هذه المصادر إما بطريقة آلية بواسطة Spidering or crawling أو بطريقة بشرية بواسطة الإنسان.[7]

الجيل الثالث من الويب Web 3.0: مصطلح لوصف مستقبل شبكة الويب العالمية حيث أنه أحد التقنيات المستحدثة في مجال تكنولوجيا المعلومات، ويعتمد علي تكنولوجيا الذكاء الأصطناعي في عمليات التصنيف والبحث وإدارة مواقع الويب، والتي تحول صفحات ومواقع الويب من مجرد مجموعة صفحات ثابتة أو صفحات ديناميكية تعتمد علي إجتماعية المعلومات إلي قواعد بيانات تقوم بفهرسة ما يتم وضعه فيها من بيانات ومن ثم التوفيق فيما بينها وبين مرادفاتها لأمكانية توزيع تلك المعلومات لأستخدامها في أكثر من سياق.[8]

---

[5] Joan M. Reitz. Availableat:<http://www.abcclio.com/ODLIS/odlis_d.aspx>.

[6] Dollar, Charles: Authentic Electronic Records strategies for long-term access, ChicagoCohasset Associates, 22 p.

[7] بن زايد عبد الرحمان.تنظيم وإسترجاع المعلومات على الشبكة العنكبوتية بين هيمنة محركات البحث وفعاليات تقنية الفاكسونومي:دراسة تحليلية.أطروحة ماجستير.-٢٠١٢م .٧٩ص.

[8] Umesha Naik, D Shivalingaiah. Comparative Study of Web 1.0, Web 2.0 and Web 3.0.- NO(2008).Available at:<http://ir.inflibnet.ac.in/handle/1944/1285>

الممسوحة ضوئيا بـ CamScanner

الويب الدلالي Semantic Web: إمتداد لشبكة العنكبوتية الحالية هكذا عرفها تيم بيرنيرز لي صاحب رؤية الويب الدلالي، وفي هذا السياق فهو تلك التطبيقات التي صممت بحيث تكون قادرة علي فهم وترميز صفحات الويب بحيث يحتوي علي معلومات دلالية تم تجهيزها خصيصاً لتفهمها برمجيات البحث والتصفح ، وهي تعتمد علي مبدأ البيانات المشتركة ، فعندما تعرف معلومة معينة يمكنك ربطها بمعلومات تتماثل مع المعلومة الأولي أوتشرحها أوتفسرها.[9] ونجد أن هناك من عرفه بأنه إطار مشترك يعمل على إتاحة تبادل البيانات وإعادة استخدامها عبر التطبيقات والمؤسسات المختلفة[10]وشبكة ويب الدلالي هي نتاج تعاون ومجهود عدد كبير من العلماء والتقنيين في مجال تقنية المعلومات وشبكة الإنترنت وذلك تحت مظلة منظمةW3C أو رابطة الشبكة العالمية World Wide Web Consortiumوالتي تعمل على إيجاد ووضع قواعد ومواصفات ومعايير للويب وتطوير تكنولوجياته الحالية.

وثائق الويب الدلالي The Semantic Web Document: هي تلك الوثائق التي تم تكويد محتواها بناء علي معايير الويب الدلالي)لغة أنطولوجيا الويب Ontology Web Languague (OWL) ، لغة التكويد قابلة الامتداد Extensible Markup Language (XML) ونماذج إطار وصف المصادر Resource Description Framework (RDF) والتي تهدف إلي نمذجة المحتوي في صورة كيانات وفئات وتوصيف العلاقات علي صعيد الفئة[11].

من خلال التعريفات السابقة إستقراءها إستقراءً جيداً إستخلصت الباحثة من جملة قراءاتها التعريف الإجرائي التالي لمصطلح:

الأرشفة الرقمية الدلالية Semantic Digital Archives : إمتداد للويب الحالي وتنطوي هذه النظم علي المهام الرئيسية لنظم الأرشفة الرقمية من إكتشاف مصادر

---

[9] Moor. M.The Semantic Web : An Introduction For Information Professionals". Thomson Reuters (Professional) Australia Limited OLC .NO(2011).

[10] Fensel, Dieter. Foundations for the Web of Information and Services : Review of 20 Years of SemanticWeb Research. - NO.(2011).Available at:< http://www.springer.com >.P5.

[11] Guha , R. & McCool , Rob & Miller. Semantic Search .- available at: http://www. Eric.org/cdrom/papers/refereed/p779/ess.html

الممسوحة ضوئيا بـ CamScanner

المعلومات ومعالجة إستفسارات المستفيدين، وترتيب النتائج البحثية وفقاً لمنطقية سؤال المستفيد ، لكن ما تمتاز به هذه النظم هو إشتمالها علي السمات التالية:

- أنها تتبني فكرة الدلالية باستخدام معايير الويب الدلالية RDF- OWL حيث أنها الخطوة القادمة من الأرشفة الرقمية.
- تعتمد علي نماذج مفاهيمية بحيث تنظر إلي الوثائق المكشفة علي أنها كيانات لها خصائص وسمات وعلاقات.

تساعد نظم الأرشفة الدلالية علي إستباط علاقات بين كلمات الإستفسار وتعزيز إمكانية التشغيل التبادلي بين الأنظمة في الميادين المختلفة لأنه يغطي الويب الدلالي- الأرشفة الرقمية-النظم الآلية في الأرشفة.

7/1 الدراسات السابقة :

١. مؤمن السيد النشرتى. محركات البحث الدلالية على شبكة الإنترنت:دراسة تحليلية لوضع مواصفات محركات بحث دلالية عربية. أطروحة (ماجستير).- جامعة القاهرة، إبريل ٢٠١٢.

هدفت الدراسة الي التعرف علي محركات البحث الدلالية من خلال حصر غالبية الخوارزميات والآليات التي تعتمد عليها محركات البحث في استرجاع المحتوي، وخوارزميات عملها وذلك لدراسة التحديات التي تواجه خوارزميات محركات البحث في استرجاع المحتوي العربي، لوضع مواصفات محركات بحث دلالية عربية.

2.Claudio Cortese and Glauco Mantegari. Extending the Digital Archives of Italian Psychology With Semantic Data .NO(2011).Available at: http://www.springer.com/

هدفت هذه الدراسة إلي استخدام الويب الدلالي من خلال نموذج أرشيف رقمي دلالي ASPI هذا المشروع يهدف إلى إنشاء مكتبة رقمية للوثائق التاريخية لعلم النفس الإيطالية وامتداده مع البيانات الدلالية، لكي يعد الويب الدلالي بوابة جديدة للبحث والأسترجاع

الممسوحة ضوئيا بـ CamScanner

والتخزين، وتوصلت الدراسة إلي أن تكنولوجيا الويب الدلالي خطوه إلي الأمام في مجال الأرشفة الرقمية.

3.Winfried Bergmeyer .The KEEP Emulation Framework. In the 1st International conference on Semantic Digital Archives (SDA 2011) , Berlin, Germany

هدفت الدراسة إلي التعرف علي أحد نظم الأرشفة الدلالية KEEP كنموذج واتضح أنه أرشيف سيخلق محاكاة تمكن الوصول إلى المواد الرقمية المخزنة بتقنيات عالية ، بحيث يمكن الوصول إليها بشكل آمن.

0/2 الويب الدلالي Semantic Web :

عرف الويب الدلالي **Semantic Web** من قبل مبتكره تيم بيرنيرز لي **Tim Berners-Lee** صاحب رؤية الويب الدلالي إمتداد لشبكة العنكبوتية الحالية، الذي تحدد فيه مفاهيم ومعاني المعلومات بوضوح لتمكين التعامل بين الآلات والبشر[١٢]، واتفق عدد من الباحثين منهم **Vadivu**، **Liyang Yu** علي مفهوم الويب الدلالي أنه هو الويب الذي يشتمل علي وثائق الويب الدلالي **Semantic Web document** (**SWD**) وتلك الوثائق متاحة علي الخط المباشر مكتوبة بلغة الويب الدلالي بصيغة إطار وصف المصدر **RDF**، أو لغة خريطة أنطولوجيا الويب **OWL** ومتاحة للوصول إليها من قبل مستخدمي البرامج الآلية[١٣]، وتشتمل علي ميتاداتا دلالية لتسجيل معلومات تفصيلية عن كل وثيقة.[١٤]

وتري الباحثة من مجمل ما ذكر أن الويب الدلالي عبارة عن تبادل البيانات من خلال أكواد وصف العلاقة بين المعلومات، ثم معالجة هذه البيانات بشكل منطقي استدلالي تحليلي، لذا

[12] Hendler, J. Web 3.0: The Dawn of Semantic Search) .No. (2010).- Accessed <24/9/2014>.- Available at:<http://www.rpi.edu/dept/IT/articles/The_Dawn_of_Semantic_Search.pdf>.p5>

[13] Vadivu, S. L.A survey on semantic web mining based web search engines. ARPN Journal of Engineering and Applied Sciences.No.(2011). Accessed <23/9/2014>.p3.

[14] Liyang Yu.A Developer's Guide to the Semantic Web).No. (2010).- Accessed <28/9/2014>.- Available at:< www.springer.com>.p3.

الممسوحة ضوئيا بـ CamScanner

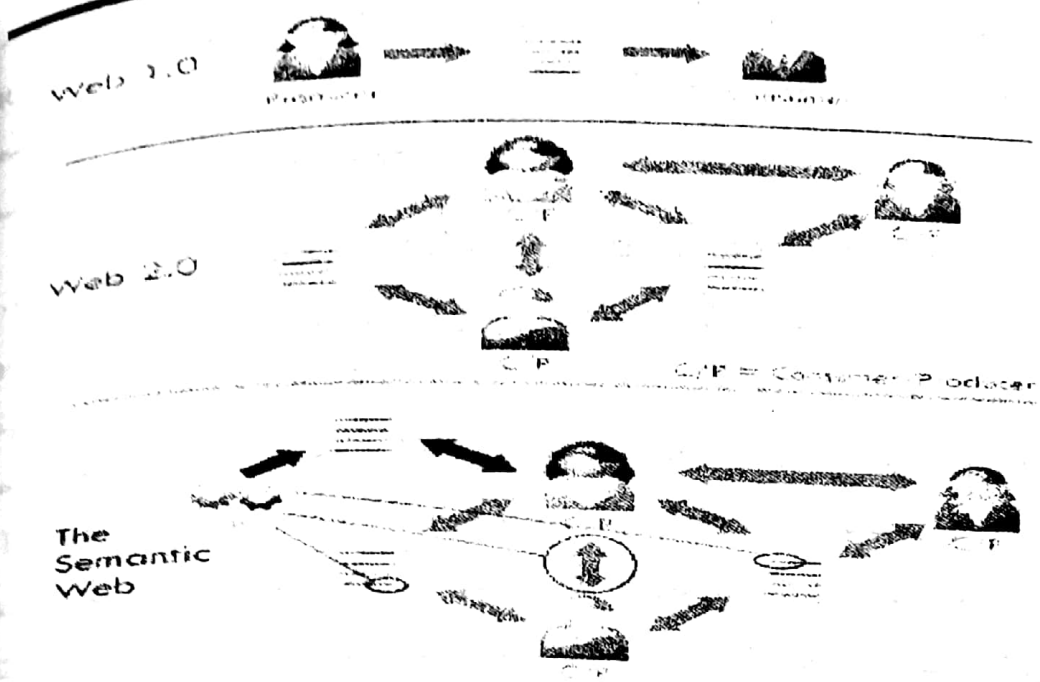
هو بسيطة وثائق متاحة علي الخط المباشر مكتوبة بلغة الويب الدلالي بصيغة إطار وصف المصدر RDF أن لغة أنطولوجيا الويب OWL وتشتمل علي ميتاداتا دلالية لتسجيل معلومات تفصيلية عن كل وثيقة .

1.2 بنية عمل الويب الدلالي *The* Architecture *Semantic* Web :

لفهم بينة الويب الدلالي لابد التعرف على خوارزمية عمل الشبكة العنكبوتية الحالية تتشكل البنى الأساسية للشبكة العنكبوتية من النصوص الفائقة HTML والتي من خلالها يتم توصيف كافة محتويات صفحات المواقع ، ويتمثل العنصر الثاني في بروتوكول HTTPوالذي طوره أيضا مخترع الويب Tim Berners Lee فلكي تعمل الشبكة العنكبوتية على الإنترنت كان لابد من تطوير آلية لعنونة الوثائق في مختلف الحاسبات تمثلت هذه الآلية في بروتكول نقل النص الفائق والذي يسمح للحاسبات بطلب الوثائق من الخوادم ، وإعتمدت آلية عنونة الوثائق والمواقع على ما يعرف بمحدد المصادر URL ، وبالرغم من ذلك عجزت النصوص الفائقة في تقديم توصيف لما تعرضة من محتوى المواقع فضلا عن عجز الأكواد الخاصة ببنية هذة الوثائق عن القيام بمشاركة الوثائق بين التطبيقات المختلفة.

ولأن عملية بناء شبكة و الويب الدلالية جارية حالياً فلابد من تحديد هيكلها ليتم بذلك ملء الهيكل أو البناء من أجل جعل هذه المهمة سهلة وممكنه ينبغي أن نبدأ بالمهام البسيطة كما هو موضح بالشكل:

الممسوحة ضوئيا بـ CamScanner



شكل رقم (1.2) يوضح تفاعلية شبكة الويب [١٥]

كما يوضح الشكل السابق أن رابطة الويب العالمية طورت بنيتها المعمارية (**Web1.0-Web2.0**) وأتضح من خلال الشكل أن الجيل الأول من الويب **Web1.0** (ويب شخصي –ويب القراءة فقط) ،الجيل الثاني من الويب **Web2.0** (ويب تفاعلي – القراءة والكتابة) ،والجيل الثالث من الويب خاص بالويب الدلالي **Semantic Web** ويساعد الويب الدلالي علي تمكين الحاسبات من:

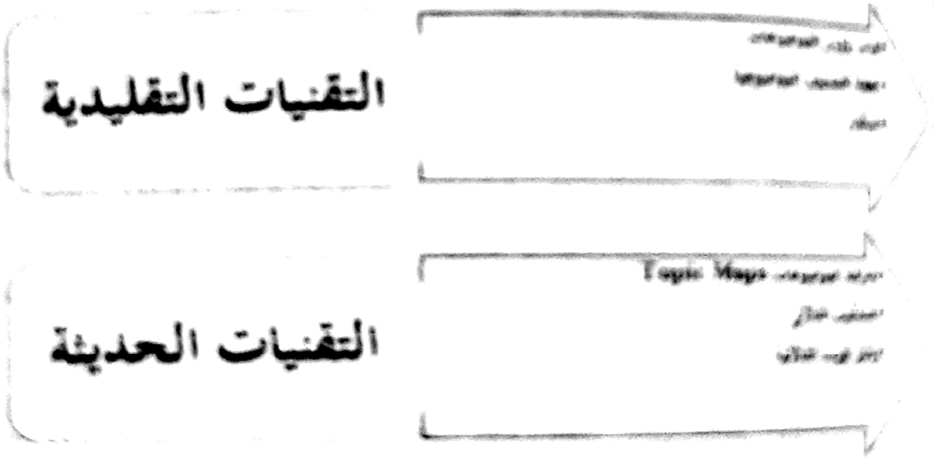
- فهم ما تعرضة من بيانات.
- توفير قابلية التشغيل المتبادل بين الانظمة المختلفة من واقع الإنطولوجيات لتوفير فهم مشترك للبيانات.
- القدرة على استخلاص معلومات جديدة مما هو متاح لديها من بيانات.

[١٥] نورة بنت ناصر الهزاني.مفهوم وبنية الويب الدلالي.-مجلة المعلوماتية.- جامعة الأميرة نورة بنت عبد الرحمن، ع٣٥(يوليو ٢٠١١). ٥٩ص.

الممسوحة ضوئيا بـ CamScanner

تقنيات البحث الدلالي:

(1) Dunsire, Gordon; Mirna Willer أحد التقنيات الحديثة التي تستخدم [illegible] الدلالي لإسترجاع المعلومات من النظام، وذلك اعتمادا على [illegible] التي ترغب المستفيد الحصول على نتائج بحوثها، فضلاً عن أن يسترجع النظام [illegible] على الترتيب التتابع للمواقع؛ فإن البحث الدلالي يكون على تقديم نتائج تتفق [illegible] (2) كما ذكر Hendler أنها المنهجية التي بمقدورها إسترجاع دلالات [illegible] بمفهومها بنوعية المعلومات المسترجعة فلا يقتصر البحث والإسترجاع على [illegible] بل يتخطى ذلك للوصول إلى معاني ودلالات الخاصة بمفردات والكلمات [illegible] في ذلك على مجموعة من التقنيات التقليدية والحديثة تعود أصولها في [illegible] المكتبات والمعلومات كالمكانز وخطط التصنيف وقوائم رؤوس الموضوعات، [illegible] الحديثة طورت في مضمار الويب الدلالي لخرائط الموضوعات Topic Maps والشكلات الدلالية والتشكيف الدلالي وونتاق الويب الدلالية ويوضح الشكل رقم (2.2) التقنيات المستخدمة في البحث والإسترجاع:



شكل رقم (2.2)يوضح التقنيات التقليدية والحديثة في البحث والإسترجاع.

(1) Dunsire, Gordon; Mirna Willer. Standard library metadata models and ontologies for the Semantic Web. No. (2011). p.- Accessed <23/12/2014> Available at:< http://dx.doi.org/10.1108/07419051111145115>.31p

(2) Hendler, J. (2010). Web 3.0: The Dawn of Semantic Search. No August 2010. Accessed <21/12/2014>.-Available at:< http://www.rpi.edu/dept/IT/articles/The_Dawn_of_Semantic_Search.pdf>.29p

الممسوحة ضوئيا بـ CamScanner

أولاً :التقنيات التقليدية :

أ. **قوائم رؤوس الموضوعات Subject Heading:** هدفت هذه الأداة إلي حصر المفاهيم والمعرفة في صورة مفاهيم لغوية ، وتنظم هذه المفاهيم عادة في نسق هجائي، وتساعد هذه الأداة علي توفير المترادافات اللغوية للمصطلح المستخدم للتعبير عن المفهوم .

ب.**نظم التصنيف الموضوعيةTaxonomy:** هو نظم التصنيف التي تعمل علي تمثيل المفاهيم والمعرفة في صورة هرمية تتداعي فيها القطاعات من الأعم إلي الخاص إلي الأكثر خصوصية،وقد تستخدم الألفاظ في التعبير عن المفاهيم أو قد تستخدم رموز أخري كالأرقام كما في الحال تصنيف ديوي العشري.

ت.**المكانز الموضوعية Thesaurus:** أحد الأدوات التي تعمل علي بناء شبكة من العلاقات بين المصطلحات والمفاهيم في مجال موضوعي معين، وتتخذ هذ العلاقات صور عديدة من الترابطية **Association** والهرمية **Hierarchical** والتكافؤية **Equivalence**.[١٨]

ثانياً: التقنيات الدلالية :

أ.**التكشيف الدلالي الكامن Latent Semantic Index (LSI) :** هي أحد أداوت التكشيف والإسترجاع التي ظهرت في غضون الويب الدلالي وتلجأ هذه التقنية إلي تكشيف السياق الخاص باالنصوص المكشفة حيث تقوم هذه التقنية بإسترجاع المحتوي المفاهيمي الوارد في النص ثم إنشاء روابط بين المصطلحات المعبرة عن المفاهيم وبين ما تشير إليه من مفاهيم، ومن جانب آخر تلجأ إلي عمل كشاف بالقيم **Values** المختلفة وربطها بالمصطلح المستخدم، كما تعمل هذه التقنية علي تحديد درجات مسبقة علي الألفاظ المستخدمة في التعبير عن المفهوم الواحد بين ما هو شائع منها في الإستخدام وما هو نادر الإستخدام ، وعلي الرغم من إتسام تقنية التكشيف الدلالي الكامن بالكثير من الميزات إلا أنها عاجزة عن تحقيق غايتها أمام الكم الهائل من مصادر المعلومات لان ما

[18] Claire Sibille,Grimoüard.ARCHIVES AND LINKED DATA: ARE OUR TOOLS READY TO "COMPLETE THPICTURE"? TWO CASE STUDIES.NO.2014).- Accessed <5/11/2014>.- www.girona.cat/web/ica2014/ponents/textos/id9.pdf.19p.

الممسوحة ضوئيا بـ CamScanner

تقوم به التقنية يتمثل في ربط المصطلحات بعضها البعض من جانب وربط المصطلحات المختلفة بالمفهوم الواحد الأمر الذي يستتبع إنشاء كشافات وقواعد بيانات ضخمة، وبالتالي يصعب تطبيق هذه التقنية في ظل ضخامه الويب وإتساع حجمه وتنوع لغات وأنواع وثائقه.[19]

ب. خرائط الموضوعات **Topic Maps**: هي مواصفة قياسية دولية صدرت عن المنظمة الدولية **ISO**، تهدف إلي تمثيل المعلومات من خلال نموذج مهيكل **Information Modelstructured** وتمثيل العلاقة بين المفاهيم ومصادر المعلومات وبنية هذا النموذج تتمثل في مستويين :

المستوي الأول/المستوي الموضوعي **Topic Space** والذي يشير إلي الموضوعات التي تمثل المفاهيم المجردة.

المستوي الثاني /مستوي مصادر المعلومات **Resource Space** والذي يتكون من مصادر المعلومات الرقمية ،ويتصل كلا المستويين ببعضهما البعض من خلال مجموعة من الروابط المحددة والموصفه، فالموضوعات ترتبط مع بعضها البعض من خلال روابط تعرف باسم **association connections** وتتصل المصادر بالمفاهيم التي تعبر عنها بنمط من الروابط يعرف باسم **occurrence connection.**

ج. وثائق الويب الدلالي **The Semantic Web Document**: هي تلك الوثائق التي تم تكويد محتواها بناء علي لغات الويب الدلالي (**OWL-XML-RDF**) والتي تهدف إلي نمذجة المحتوي في صورة كيانات وفئات وتوصيف العلاقات علي صعيد الفئة[20].

[19] Barbara Rosario .Latent Semantic Indexing: An overview .NO.2001).- Accessed <7/11/2014>.- www.cse.msu.edu/~cse960/Papers/LSI/LSI.pdf.2001. 13p.

[20] Guha , R. & McCool , Rob & Miller. Semantic Search.- Accessed <5/11/2014>.-available at: <http://www. Eric.org/cdrom/papers/refereed/p779/ess.html>8p

الممسوحة ضوئيا بـ CamScanner

كما يمكن تحديد خصائص البحث الدلالي من خلال مقارنته بالبحث التقليدي ويمكن حصرهما على النحو الآتي:

الفئة الأولى: البحث المعتمد على الألفاظ.

الفئة الثانية: البحث المعتمد على المفاهيم. ويمكن مقارنتهم كالتالي:

أوجة المقارنة	البحث المعتمد على الألفاظ[21] (البحث التقليدي)	البحث المعتمد على المفاهيم[22] (البحث الدلالي)
تعريفه	بحث يعتمد على المضاهاه المعجمية واللغوية في البحث والأسترجاع.	هو ذلك البحث الذي يعتمد على الأسترجاع وفقا للمفاهيم والسياق الدلالي للوثائق.
إعتماده	الأعتماد على الكلمات المفتاحية في كل من الأستفسار والتكشيف.	البحث ذات الصله أي تقوم اداه البحث باقتراح بعض الموضوعات ذات الصلة بالأستفسار المقدم من قبل المستفيد.
عرض الوثائق	العرض دون الفهم:عرض الوثائق بصورة قياسية دون أية محاولة لأستثمارها او تكامل محتواها مع برمجيات أخري.	البحث وفقاً للنص الكامل أي التخلي عن الأسترجاع وفقا للكلمات المفتاحية والأعتماد على السياق ومفهوم النص.
البحث	الوثائق تسترجع دون أي ربط معرفي بينها وبين نظائرها.	القدرة على توفير البحث وفقا للمشتقات اللغوية لمصطلحات البحث.
الإسترجاع	يعتمد علي تمثيل الوثائق والبحث عنها وإسترجاعها من خلال تكويد المحتوي بلغةHTML المحدودة.	الإسترجاع المفاهيمي الذي يتمثل في القدرة علي إسترجاع الوثائق لا ترد فيها ألفاظ الإستفسار لكن تعالج المفهوم.

جدول رقم (1.2)يوضح الفروق بين البحث التقليدي والبحث الدلالي.

[21] David Green. the semantic web: search engines, xml, and intelligent agents.IEEE. .-No(2007).- Accessed <5/11/2014>.- Available at:< http:www.davidgreen.me.uk>.3p

[22] Kowalski, G). Information retrieval architecture and algorithms. New York .-No(2011).-).- Accessed <1/11/2014>.- Available at:<http://www. Springer.org/>.7p.

الممسوحة ضوئيا بـ CamScanner

الدراسة المقارنة بين محرك البحث جوجل GOOGLE ونظام أرشفة

الدراسة
الدراسة الحالي Vivo:
أولاً محرك البحث GOOGLE
https://www.google.com.eg/

١ نشأته

محرك البحث (Google)أكثر محركات البحث شهرة على المستوى العالمي، نظراً لما يتمتع به من سرعة في ملاحقة الصفحات الجديدة وتكشيفها، حيث يبلغ حجم قاعدة بياناته وفقا لآخر التقديرات حوالي *9 مليار صفحة*وكما هو معروف فقد قام بتطوير هذا المحرك كل من طلبة الدكتوراه بجامعة ستانفورد عام *1998 مما* Larry Page و Sergey Brin، والمصطلح جوجل يعنى بالإنجليزية ملايين المليارات كما تعنى أيضا يبحلق أو ينظر بشدة وتركيز مما يعكس رغبة مؤسسيه على ملاحقة وتكشيف أكبر قدر ممكن من صفحات ومواقع الويب، كما يعكس أيضا إدراكهم منذ البداية للنمو السريع لشبكة الإنترنت، وتشير إحصائيات محرك البحث جوجل إلى أنه يجيب على أكثر من ١٥٠ مليون استفسار يوميا بمختلف لغات العالم، حيث يتيح البحث بحوالي ٣٥ لغة منها اللغة العربية .[٢٣]

٢ خصائص محرك البحث جوجل GOOGLE وملامح بحثه:

و أثبت محرك البحث جوجل كفاءة كبيرة في تعامله مع اللغة العربية من خلال العديد من التجارب والدراسات التي سعت إلى تقييم أداء محركات البحث ،ومن أهم الملامح التي تميز محرك البحث جوجل هو

- قدرته الفائقة على عرض وترتيب النتائج وفقا لارتباطها بموضوع الاستفسار ، ويرجع الفضل في هذه الميزة إلى الطريقة التي يعتمد عليها محرك البحث في ترتيب نتائج الاسترجاع والتي تعرف بنموذج بترتيب الصفحة Page Rank ، وفي معدلات

[23] **Jakobsson Peter, Stiernstedt Fredrik. The politics of organizing information on the web: computing centers and natural languages** .Available At:<www.web.mit.edu/commforum/ mit6/papers/Jakobsson.pdf>

الممسوحة ضوئيا بـ CamScanner

الاستدعاء Recall والتحقيق Precision ودقة الترتيب Ranking Accuracy.

- يجمع Google بين Page Rank وتقنيات مطابقة النص Text Matching المعقدة ليجد صفحات مهمة وتلائم موضوع البحث على السواء، ولا يتوقف Google عند عدد المرات التي تظهر فيها عباره معينة في الصفحة، بل يفحص كل أوجه محتويات الصفحة ومحتويات الصفحات المرتبطة بها ليعرف ما إذا كانت مطابقة للبحث أم لا.[٢٤]
- يرى أندريه روكس (Andre Roux) معد الكتاب السنوي لمحركات البحث Search Engines Yearbook أنه توجد مجموعة من المبررات المنطقية التي تدفع المستفيدين إلى تفضيل جوجل عن غيره من محركات البحث وزيادة معدلات استخدامه من جانب رواد شبكة الإنترنت، وتشمل هذه المبررات ما يلي :

- التغطية في جوجل Google Coverage: أن محرك البحث جوجل أكبر محركات البحث من حيث تغطيته لصفحات ومواقع الويب كما انه أسرع محركات البحث المتاحة على الشبكة العنكبوتية.

- بنية محرك البحث Search Engine Structure: الأساس الذي أعتمد عليه مصممي جوجل هو أساس ديمقراطي إلى حد كبير ويتناسب مع طبيعة الشبكة العنكبوتية بما تتميز به من روابط فائقة حتى أن مصمم الشبكة العنكبوتية تيم برنر لى يرى أن مصممي جوجل تفوقوا على كل محركات البحث الأخرى بإدراكهم هذه الميزة التي جعلت من جوجل بيئة متميزة لاسترجاع المعلومات من الشبكة العنكبوتية.

- الحياد Neutral and Un-Bias: إن محرك البحث جوجل يتميز أيضا بأنه يسعى لتنظيم شبكة الإنترنت بحيادية تامة حيث يقوم بعرض النتائج مدفوعة الأجر في موقع منفصل وبعيد تماماً عن نتائج البحث الأساسية.

[24] Wall, Aaron. Search Marketing. History of Search Engines & Web History. .-No(2005).-).- Accessed <1/11/2014>.- Available At: <.http://www.search-marketing.info/search-engine-history >

الممسوحة ضوئيا بـ CamScanner

- دقة النتائج المسترجعة Accuracy of Retrieved Results: تتميز النتائج المسترجعة من محرك البحث جوجل بالدقة المتناهية حيث يعتمد على أسلوب المطابقة الحرفية Exact Match بالتالي فهو لا يسترجع إلا النتائج التي تتطابق مع كلمات البحث سواء ضمن نص الصفحة أو ضمن الارتباطات التي تشير إلى الصفحة.

- وصف النتائج Results Description: يقدم جوجل وغيره من محركات البحث وصف دقيق للصفحة المسترجعة في صورة مقتطفات من النص الذي يطابق البحث، وبالطبع تتغير هذه المقتطفات وفقا لكلمات البحث. [٢٥]

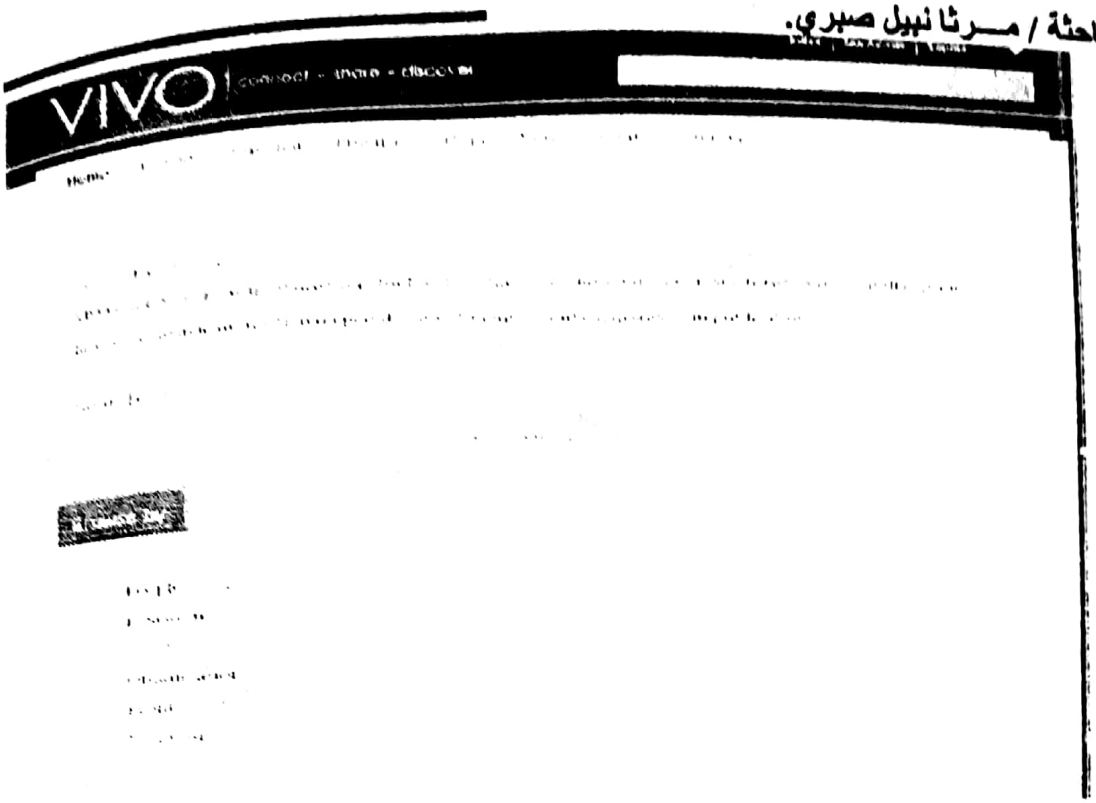
2/3 ثانياً : نظام VIVO:
http://www.vivoweb.org/

١. نشأته:

طور هذا النظام من جانب فريق علمي يضم جامعة كورنيل وأنديانا وجامعة ولاية فلوريدا ومعهد مكربيس للأبحاث The Cornell University, the University of Florida, The Scripps Research Institute في المملكة المتحدة عام 2009 ،كأداه لإلتقاط وتحليل وتكشيف البيانات لمستخدمي الويب الدلالي وتطبيقاته، حيث يمثل هذ النظام برنامج مفتوح المصدر Open-Source Software ،حيث أعلن فريق العمل أن هذا النظام يمثل جيلاً جديداً من نظم الارشفة الدلالية نظراً لقدرته علي توفير البيانات الدلالية من أنطولوجيات وكيانات وعلاقات دلالية لبرمجيات الويب الدلالي المختلفة Semantic WEB Application بصورة آنية ومتكاملة، يجري حالياً تطوير نظام VIVO كمشروع مفتوح المصدر بمشاركة المجتمع من جميع أنحاء العالم ، ويوضح الشكل رقم (2.3) واجهة التعامل لنظام الأرشفة الدلالي VIVO.

[٢٥] بن زايد عبد الرحمان.المرجع السابق .١٦٣ص

الممسوحة ضوئيا بـ CamScanner



شكل رقم (2.3) يوضح واجهة التعامل لنظام الأرشفة الدلالي VIVO

٢. خصائص النظام :

يتميز هذا النظام أنه قام بإعداه فريق عمل لحصر متطلبات برمجيات الويب الدلالي التي تعتمد علي وثائق الويب الدلالي من أنطولوجيا وغيرها ***Semantic Web Application*** ،وقد أسفرت دراستهم أن برمجيات الويب الدلالي أخدت نهجاً جديداً في الإعتماد علي الإنطولوجيات ،فعوضاً عن اعتماد علي أنطولوجيا واحدة فقط في عملياتها ،أصبحت في الوقت الراهن بحاجة إلبي الجمع بين الإنطولوجيات المختلفة في وقت التشغيل الآني لها، حيث وفرت الحاجة إلي توفير آداه توفر الإسترجاع الآلي لكافة وحدات المعرفة ذات الصلة.

وتتمثل أبرز خصائصه وراء نشأة نظام VIVO في التالي :

- دراسة بنية وهيكل الويب الدلالي بصوة فعاله ، ودعم البرمجيات التي تدعم وثائق الويب الدلالي وما بها من أنطولوجيا بحيث يكون بمثابة مستودع للأنطولوجيات المختلفة مما يمكن البرمجيات من إعادة إستخدامها.

الممسوحة ضوئيا بـ CamScanner

- قصور نظم الأرشفة التقليدية علي تحقيق الإسترجاع المفاهيمي من خلال الوثائق التقليدية أو وثائق الويب الدلالي ، وتم تطويره بحيث يعمل علي إكتشاف والتقاط أكبر قدر من وثائق الويب الدلالي وتجميعها وتكشيفها.

٣. ملامح البحث:

تدعم إمكانيات الوصول الدلالي للإنطولوجيات :يعد هذا المبدأ من اهم المبادئ البرمجيات فنظراً لتنوع متطلبات الويب الدلالي وتطبيقاته بين مستويات مختلفة في الوصول إلي الإنطولوجيات فمنهم ما يعتمد علي علي المصطلحات ومنهم ما يعتمد علي العلاقات ومنهم ما يعتمد علي الميتاداتا الخاصة بالإنطولوجيا ،الأمر الذي يستلزم توفير أداه بحث توفر نقاط تتسم بالدلالية لفهم مستوي كل برنامج علي حده.[٢٦]كما يمكن النظام من حصر العلاقات بين الإنطولوجيات ،كما يدعم متطلبات التشغيل البيني بين برمجيات الويب الدلالي.

توفير نتائج بحث دلالية منتقاه من وثائق الويب الدلالي في صورة تمكن البشر من التعامل معها بسهولة ويسر في ظل علي أن تكود هذه النتائج إعتماداً علي لغات النص الفائق **HTML**، ويوضح الشكل التالي واجهه التعامل في نظام **VIVO** وتدعيم نتائج بحثة للغات النص الفائق بالأضافة إليRDF، وبتطبيق البحث في علي نظام VIVO الدلالي إتضح البحث والإسترجاع وفقاً للغات التكويد **OWL& RDF** التي تهدف إلي إسترجاع المعلومات وفقاً لهذه اللغات أو لنماذج الإسترجاع في موضوع محدد .

كما يوضح أن صيغ الإستفسار في نظام **VIVO** تعتمد علي الكلمات المفتاحية شأنها نظم الأرشفة التقليدية ،كما تسمح بتقديم الإستفسار وفقاً لأسماء القطاعات الموضوعية **Classes** أو صيغ التكويد **RDF**.

[26] Jon Corson-Rikert, Stella Mitchell, and Brian Lowe .The VIVO.- Cornell University , University of Florida.- Accessed <9/1/2015>.-NO (2012).Available at: <http://vivoweb.org/blog/2012/10/vivo.pdf>

الممسوحة ضوئيا بـ CamScanner

يدعم النظام البحث والإسترجاع من خلال URL بحيث يوجه المستفيد إستفساره عن موقع Locations والتي تعتمد علي إسترجاع الوثيقة وفقاً ل URL.

وبمقارنة نتائج بحث نظم الأرشفة الرقمية الدلالية SDAS والتي يمثلها نظام VIVOالدلالي مع محرك البحث GOOGLE بطريقة تطبيقية عملية وعرضها علي أعضاء هيئة التدريس بكلية الطب إستخدمت الباحثة 4 مصطلحات بحث مختلفة وهي : , "Recombinant" ,"CTSI", 'Computational biology", "Conlon Michael" ،وقد حرصت علي تنوع المصطلحات لتشمل إسم شخصية ومصطلح طبي ومؤسسة ثقافية ،والغرض من تنوع المصطلحات هو حرص الباحثة علي التعرف علي الخصائص المختلفة والمميزات التي يمنحها نظام الأرشفة الدلالي Vivo مقارنته بجوجل GOOGLE وتبين من عملية الإسترجاع الآتي:

١. جاء عرض النتائج في نظام الأرشيف الدلالي Vivoبشكل متكامل حيث تظهر نتيجه البحث مقسمة إلي فئات وهي كالاتي الويب ،منظمات،الأخبار،الاشخاص وقد يضاف إليها نتائج آخري ذات صله حول موضوعات آخري، أما في محرك البحث جوجل يتطلب من المستفيد إختيار الفئة التي يرغب البحث فيها ومن ذلك الويب، الصور، الأخبار، أوإجابات حول الموضوع .

٢. أن نتائج البحث في Vivo جاء أقل من نتائج محرك البحث GOOGLE من حيث العدد، والسبب في ذلك يرجع إلي أن نظام Vivo الدلالي الطبي يعمل علي إنتقاء بناءً علي الروابط الدلالية وكذلك تركيز نظام Vivo أكثر حول موضوع الإستفسار وما يتعلق به، سواء للمواد النصية أو المصور أو المسجلة فعلي سبيل المثال بلغت عدد نتائج البحث عن مصطلح "Recombinant" مقسمة علي فئات كالتالي 238أشخاص ،492 صفحات منشورة، 61 منظمات، 48 أنشطة ،أما عند البحث عن المصطلح في محرك البحث جوجل بلغ عدد النتائج المسترجعة الويب 21,000,000 الأخبار 14,600, الفيديو77,400 التطبيقات 2,540 تبين من خلال البحث أن نظام أرشفة Vivo يعطي نتائج قليلة ومركزة مقارنةً بمحرك بحث جوجل GOOGLE.

الممسوحة ضوئيا بـ CamScanner

٣. عند الإسترجاع علي مصطلح "Computational biology" في نظام أرشفة VIVO إتضح انه تم تقسيم نتائج البحث إلي قسم Biology في الطب وProgram ومؤسسات Association وغيرها لذا فإقتصرت النتائج في نظام VIVO حول البيولوجية الحاسوبية فقط حيث تم مراعاه السياق الذي جاءت فيه الكلمتين ،أما نتائج البحث في جوجل تضمنت فقط المصطلحات الواردة ضمن الإستفسار.

٤. عند إسترجاع نتائج البحث عن مؤسسة طبية مثل "CTSI" إتضح أنه في VIVO تم توضيح بوضوح والأبحاث الموجودة علي الموقع واللجنة المؤسسة ،ومما سبق إتضح أن الويب الدلالي يعطي نتائج قليلة عن موضوع البحث لكنها أكثر إرتباطاً بإستفسار المستفيد بالمقارنة بمحرك البحث جوجل GOOGLE

٥. وعند البحث عن مصطلح يتضمن إسم شخص "Conlon Michael" يعطي نظام أرشفة VIVO معلومات محددة عن الشخص المراد البحث ،عنه بعكس محرك البحث جوجل GOOGLE تضمنت نتائج بمطابقة الكلمات المدخلة مع كلمات البحث.

٦. حاولت الباحثة إجراء بحث باللغة العربية في نظام الأرشفة الدلالي VIVO ،وذلك لمقارنة نتائج البحث باللغة العربية في النظام،ولكن تبين أن النظام لا يتعامل مع اللغة العربية مطلقاً فأحيانا لا تسترجع عملية البحث أي نتيجه وفي أحيانا أخري تتضمن عملية الإسترجاع علي النظام أنه لم يتم الإسترجاع باللغة العربية ويعطيني كلمات مقترحة للإسترجاع باللغة الإنجليزية .

الممسوحة ضوئيا بـ CamScanner

***نتائج الدراسة**

وعند إجراء البحث علي محرك البحث جوجل(متقدم) ومقارنة نتائج البحث مع نظم الأرشفة الدلالية VIVO إتضح التالي:

تختلف النظم الدلالية عن نظريتها التقليدية من جوانب عدة،ما يتعلق بتقنيات تطوير تلك النظم، وكذلك آلية عملها والنتائج التي تقدمها للمستفيد ....إلخ، وفيما يلي يوضح ما يميز نظم الأرشفة الدلالية عن محركات البحث التقليدية علي النحو التالي:

١. يقدم البحث الدلالي نتائج أكثر صلة وارتباطًا باستفسارات المستفيدين، وذلك لاعتماده على دلالات مصطلحات البحث بدلاً من الاعتماد على الترتيب الشائع للمواقع.

٢. يعمل نظام أرشفة VIVO علي ربط العلاقات بين الوثائق والأشخاص والأماكن والأحداث وبذلك فإن نتائج البحث التقليدي تأخذ شكل قائمة بالوثائق وصفحات الويب،وهذا يعني أن نتائج البحث الدلالي لا تقتصر فقط علي قائمة بالوثائق أو صفحات الويب فقط بلكنها تتجاوز ذلك إلي تقديم البيانات ذات الصلة(المؤسسة التابع لها أ، الموضوع أ،الموقع أ،الوثيقة).

٣. يعتمد البحث الدلالي في عمله علي تطبيقات الويب الدلالي من أبرزها صيغ XML,OWL,PDF.

٤. لا يتناسب البحث الدلالي مع طريقة البحث الملاحي Navigational Search التي يحاول فيها المستفيد عادة إسترجاع وثيقة معينة وبالتالي فإنه يستخدم كلمات متفرقة بدون معني لضمان ورودها ضمن الوثيقة التي يتم إسترجاعها،وفي المقابل فإن البحث الدلالي يستخدم فيه المستفيد كلمة ذات معني للوصول إلي نتائج حولها، وقد تكون تلك الكلمة إسم شخص أو مكان.

الممسوحة ضوئيا بـ CamScanner

** توصيات الدراسة :

١. تدريب الطلاب والباحثين على إستخدام الويب الدلالي Semantic Web .
٢. تبني الدراسة وإستخدام تطبيقات الويب الدلالي عامة ،ونظم الأرشفة الدلالية خاصة VIVO في المؤسسات الطبية.
٣. الحاجة لتكثيف الدورات في تطبيقات الويب الدلالي Semantic Web ، بحيث تتاح لأعضاء هيئة التدريس بكليات الطب المهارات الكفيلة بالاستخدام الفعال للإنترنت في التدريس.
٤. تطوير نظم أرشفة دلالية لإسترجاع المعلومات باللغة العربية.

الممسوحة ضوئيا بـ CamScanner