مجلة جامعة تشرين للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (37) العدد (6) 2015
Tishreen University Journal for Research and Scientific Studies - Engineering Sciences Series Vol. (37) No. (6) 2015

استكشاف المعارف في الويب الدلالي
(استرجاع المعلومات من قواعد المعرفة)

الدكتور باسل الخطيب*
كامل جعفر**

(تاريخ الإيداع 5 / 7 / 2015. قُبِل للنشر في 17/ 11 / 2015)

□ ملخّص □

يُعتبر الويب الدلالي ثورة جديدة في عالم الويب، حيث تصبح المعلومات والبيانات قابلة للمعالجة منطقياً من قِبل برامج الحاسوب، بحيث تتحول تلك المعلومات والبيانات إلى شبكة معارف ذات معنى مُحدّد. يُشكّل الويب الدلالي مستقبل الويب الحالي مع الإشارة إلى أن البحوث والدراسات العربية مازالت قليلة نسبياً في هذا المجال. نعرض في هذه الورقة دراسة مرجعية عن الويب الدلالي والطرق المختلفة لاستكشاف المعارف واستنباط المعلومات المفيدة من الكم الهائل من البيانات التي يوفرها الويب. كما نُقدّم مثال برمجي قمنا بتطويره لتطبيق لبعض التقانات التي يُقدمها الويب الدلالي ولاسيما طرق استكشاف المعارف فيه. نوفر في هذا المثال البرمجي المبسط بعض الخدمات المتعلقة بالتعليم العالي الحكومي السوري، مثل معلومات عن الجامعات الحكومية السورية كاسم الجامعة (الافتراضية، دمشق، تشرين، حلب والبعث)، موقع الجامعة، عنوان موقعها على شبكة الويب، عدد الطلاب فيها، وملخص عن الجامعة، مما يُساعد العملاء الأذكياء على إيجاد تلك الخدمات ديناميكيا.

الكلمات المفتاحية: الويب الدلالي – استكشاف المعارف – الانطولوجيا – التنقيب في الويب.

* **رئيس قسم الذكاء الصنعي–كلية المعلوماتية–جامعة دمشق–مدير برنامج ماجستير علوم الوب في الجامعة الافتراضية السورية– حاصل على شهادة الدكتوراه في علوم الكمبيوتر من جامعة بوردو–فرنسا– 1993.**
** **ماجستير في علوم الويب– إجازة في تكنولوجيا معلومات– الجامعة الافتراضية السورية.**

مجلة جامعة تشرين للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (37) العدد (6) 2015
Tishreen University Journal for Research and Scientific Studies - Engineering Sciences Series Vol. (37) No. (6) 2015

Knowledge Discovery in Semantic Web (Information Retrieval from Knowledge Bases)

Dr. Bassel AlKhatib*
Camel Jafar**

(Received 5 / 7 / 2015. Accepted 17 / 11 / 2015)

□ ABSTRACT □

Semantic Web is a new revolution in the world of the Web, where information and data become viable for logical processing by computer programs. Where they are transformed into meaningful data network. Although Semantic Web is considered the future of World Wide Web, the Arabic research and studies are still relatively rare in this field. Therefore, this paper gives a reference study of Semantic Web and the different methods to explore the knowledge and discover useful information from the vast amount of data provided by the web. It gives a programming example like application of some of these techniques provided by the Semantic Web and methods to discover the knowledge of it. This simplified programming example provides services related to higher education Syrian government, such as information about the Syrian public universities like the name of the university (Syrian Virtual University, Tishreen, Aleppo, Damascus, and Al Baath), address of the university, its web site, number of students and a summary of the university, which helps intelligent agents to find those services dynamically.

Keywords: Semantic Web, Web Mining, Ontology, Knowledge Discovery.

* **the web science master director at the Syrian Virtual University and the head of Artificial Intelligence department at Information Technology Faculty at Damascus University. He holds PhD degree in computer science from the University of Bordeaux-France, 1993.**
** **Master in Web Sciences- Bachelor in Information Technology - Syrian Virtual University.**

مقدمة:

كان لدينا في البداية المواقع الثابتة المصممة بلغة HTML تحت مسمّى WEB1. ثمّ أتت برامج إدارة المحتوى التي غيرت المفهوم التقليدي للو يب للمفهوم الديناميكي من حيث ربط الملفات بقواعد البيانات وإمكانية التعديل من خلال لوحات تحكّم البرامج والمنصات فظهر من خلالها مصطلح جديد وهو WEB1.5 ولكنه تطوّر وأصبح WEB2 الذي يسمح بدمج المحتويات مثل XML و Web Services. تحول مفهوم الانترنت من مصدر للمعلومات إلى مصنع للمعلومات التفاعلية، من خلال المجتمعات الافتراضية والخدمات المستضافة التي وفرت قدر عالي من التفاعلية مع المستخدم. إن ظهور الويب الدلالي"Semantic Web" كان نتيجة لرؤية (Berners-Lee *et al.* 2001) لما سيكون عليه شكل الويب في المستقبل، بأن تكون الروابط أكثر دلالية للعلاقة فيما بينها كأن تشير إلى أن هذا الرابط هو" نوع من" أو "جزء من"، أي أن يكون أكثر ذكاءً بطريقة تجعل الآلة أيضاً تفهم ماذا تعني محتويات صفحة ما في الويب وماذا تعني الروابط في تلك الصفحة.

أهمية البحث وأهدافه:

تأتي أهمية البحث من أهمية تطبيق تقنيات الويب الدلالي والتي تعمل على تمثيل محتوى الويب بشكل قابل للمعالجة من قبل الآلات واستخدام تقنيات ذكية للاستفادة من هذا التمثيل . والهدف النهائي للويب الدلالي هو الوصول بشبكة الويب العالمية لشكل يجعلها أقرب إلى قاعدة بيانات عالمية موحدة مترابطة ذات نسيج موحد تترابط فيه المعلومات المتاحة فيها مع البرمجيات وأدوات البحث كما هو الحال في قواعد البيانات العادية، وهذا ما عملنا عليه من خلال تصميم ملفات RDF تحتوي معلومات عن بعض الجامعات السورية والربط بينها بالعلاقات الممكنة من خلال تصميم انطولوجيا من نوع (RDF-Based Graph).

1. مفهوم الويب الدلالي:

تعتبر شبكة الويب أغنى مصادر للمعلومات بما تحويه من مستندات وملفات منوعة يُمكن الوصول إليها عن طريق محركات البحث التقليدية. غير أن تنظيم هذه الملفات بصورة تسهل عملية البحث فيها والوصول إليها، يعتبر أمراً غاية في الصعوبة. يضاف إلى ذلك، أنه في ظل التزايد المستمر في حجم المعلومات المنشورة في شبكة الويب أصبح من الصعوبة بمكان قيام محركات البحث بإيجاد المعلومات المناسبة.

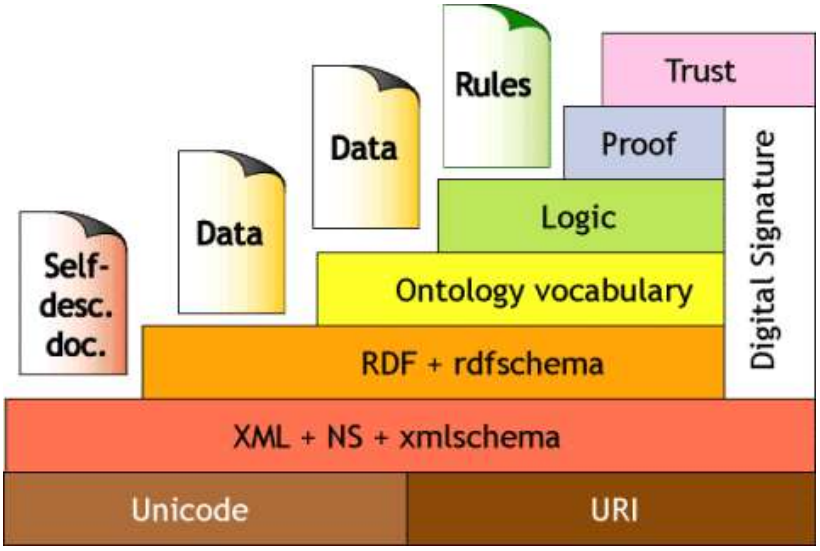
ومن هذه المشكلة ظهرت فكرة "الويب ذو الدلالة والمعاني اللفظية"، أو ما يطلق عليه بالإنجليزية مصطلح (SEMANTIC WEB) والتي هي امتداد للويب الحالية ولكن تختلف عنها بأنها تتفهم مدلولات الألفاظ والمعاني البشرية.

وانصبت جهود الباحثين في تطوير مشروع الويب الدلالي تحت إشراف منظمة (W3C) لنقل الفضاء الشبكي من مجرد مستودع ضخم للمعلومات المبعثرة إلى بناء منظم. تكون المعلومات فيه مترابطة جيداً ومفهومة من قبل الآلة بشكل يمكن لها معه إدراك العلاقات الترابطية بين عناصر المعلومة جاعلين من البحث عملية تقوم البرامج بجزء كبير منها، فنحصل على شبكة من البيانات التي يمكن معالجتها من قبل الآلات بشكل مباشر أو غير مباشر. ليصبح الويب أكثر قابليةً للفهم من قبل الآلات بواسطة بناء بنية تحتية مناسبة للعملاء الأذكياء (INTELLIGENT AGENTS) للقيام بعمليات معقدة لمستخدميهم، منها دمج المعلومات بطريقة ذكية، وتوفير الوصول المعنوي الدلالي إلى الإنترنت.[1]

2. بنية الويب الدلالي:

اقترح Berners Lee بناء طبقات للشبكة الدلالية على النحو المبين في الشكل (1):

[Berners-Lee *et al.* 2001]



الشكل (1) يبين طبقات شبكة الويب الدلالية

من خلال الشكل(1) السابق يتضح أن الويب الدلالي يتكون من نماذج بيانات تستخدم عدداً من التقنيات لتمثيلها كما هو واضح من خلال الطبقات المتعددة في النموذج وهي على النحو التالي:

1. **معرفات الموارد الموحدة Unicode , Uniform Resource Identifier** واختصارها URI وهو أسلوب التخاطب فيما بين متصفح الويب الدلالي الذي يستعمله المستخدم النهائي والمحتوى الموجود على الويب الدلالية، ومعرفات الموارد ينظر أليها على أنها شيفرات وألفاظ أو معرفات مشتركة للأشياء التي سنتفاهم بشأنها وقد تكون هذه المعرفات لفظية أو شكلية (الحروف، الأرقام، الصور) ومن أمثلة هذه المعرفات الأكثر انتشارا في العالم هو URL.

2. **لغة الترميز القابلة للتوسع Extensible Markup Language (XML):** هذه الطبقة مسؤولة عن التأكد من سلامة القواعد النحوية المستخدمة في لغات توصيف الانطولوجي في الويب الدلالي، ومطابقتها لقواعد XML Schema.

3. **لغة وصف المصدر Resource Description Framework (RDF):** توفر هذه الطبقة نموذج بيانات خاص بالويب الدلالي حيث تصف الـ RDF مصادر الويب ككائنات، وتحدد العلاقات فيما بين هذه الكائنات التي قد تكون مصادر أخرى على الويب. وهي لغة لتمثيل البيانات المشتركة التي يمكن استخدامها لتبادل وتكامل المعلومات المقروءة آلياً.

4. **مخططات العلاقات RDF Schema (RDFS):** توفر هذه الطبقة لغة لتوصيف نموذج بيانات الأنطولوجي حيث تعرف فئة الكائنات، فئة الخصائص وتعرف التسلسل الهرمي بين الكائنات والتسلسل الهرمي بين الخصائص، وتحدد مجال ومدى الخصائص.

5. **لغة أنطولوجيا الويب(OWL) Ontology Web Language:** وتوفر هذه الطبقة نموذجاً آخر من لغة توصيف الأنطولوجي أكثراً تطوراً من RDFS. تعرف أيضاً فئات الكائنات وفئات الخصائص وتعرف التسلسل الهرمي بين فئات الخصائص والكائنات، إلا أنها تحدد أيضاً نوع العلاقات بين فئات الكائنات نفسها، فيما إذا كانت مترابطة أو مستقلة عن بعضها البعض، وفيما إذا كانت العلاقات متناظرة أو متعدية أو قابلة للعد. تعرّف الـ OWL باستخدام صيغة ملفات وقواعد XML.

6. **القواعد Rules:** وتهدف هذه الطبقة لدعم قواعد استدلالية تسمح بالاستعلام والتصفية . تستخدم اللغتان السابق ذكرهما ولغات أخرى مبنية عليها لإعطاء نتائج منطقية تماماً كما يفكر البشر ولكن حتى الآن، لا يوجد لغة موصى بها لهذه الطبقة إلا أن Rule Interchange Format (RIF) هي اللغة الأكثر استخداماً.

7. **الطبقة المنطقية Logic Layer:** توفر هذه الطبقة تسهيلات لكتابة قواعد منطقية في المستندات، بحيث توفر قواعد لاستنتاج نوع ملف من ملف آخر. وتشمل هذه الطبقة على توابع منطقية تسهل عملية الاستنتاج. وتعتبر صيغة تبادل المعرفة Knowledge Interchange Format (KIF) هي اللغة المستخدمة لتحديد المنطقية في هذه الطبقة.

8. **الإثبات Proof Layer:** طبقة تعمل على التأكد من الخطوات المتخذة في توليد معرفة جديدة قد تمت بشكل مناسب وصحيح مثل التأكد من وجود المحددات، ونطاقات الأسماء، وانتماء المصطلحات إلى فئاتها التصنيفية.

9. **المصداقية Trust Layer:** تضمن هذه الطبقة أن يكون مصدر المعلومات حقيقي وأصلي، وفي الوقت نفسه تضمن ألا تصل المعلومات إلا لعملاء البرمجيات والمستخدمين المخول لهم بذلك. و يأتي تحقيق المصداقية من اعتماد التوقيع الإلكتروني والمعتمد على التشفير في المستندات.[2]

1. الأنطولوجي Ontology:

هي طريقة لتمثيل المفاهيم وذلك عن طريق الربط بينها بعلاقات ذات معنى. فالأنطولوجيا هي توصيف صوري (FORMAL)لمجال معين (DOMAIN) من خلال تحديد المفاهيم الخاصة به (CONCEPTS) و صفاتها (ATTRIBUTES) والعلاقات بين هذه المفاهيم (RELATIONS) . تكون المفاهيم عادة منظمة بشكل هرمي (HIERARCHY) بحسب علاقة التعميم والتخصيص (GENERALIZATION/SPECIALIZATION) بين هذه المفاهيم. في هذه الهرمية يكون المفهوم الخاص ابناً للمفهوم الأعم منه. مثلاً: "الطالب" هو "إنسان" و "الإنسان" هو "كائن حي" (المفهوم "طالب" هو ابن المفهوم "إنسان" والذي بدوره ابن المفهوم "كائن حي").[3]

1. الهدف الرئيسي من الأنطولوجي:

1 تكوين فهم مشترك حول بنية المعلومات يدركه كل من الإنسان والآلة.

2 جعل مجال المعرفة قابل لإعادة الاستخدام.

3 جعل فرضيات المجال واضحة ومفهومة.

4 فصل مجال المعرفة عن مجال الإجرائيات.

5 تسهيل عملية تحليل مجال المعرفة.

2. تصنيف الأنطولوجي:

1. **الأنطولوجي العامة (Generic Ontology):** وتغطي المعرفة العامة، أي أنها متعلقة بمجالات متنوعة وليس فقط متخصصة في مجال واحد، وبالتالي يمكن تطبيقها على مجموعة مختلفة من أنواع المجالات.

2. **الأنطولوجي الخاصة بمجال (Domain Ontology)**: تغطي المعرفة في مجال محدد وتكون مشتركة بين مختصي هذا المجال.

3. **أنطولوجي التطبيق (Application ontology)**: تغطي المعرفة الضرورية لتطبيق معين ويقتصر مستخدمون هذه الأنطولوجي على مطوري هذا التطبيق. [4]

4. استكشاف المعرفة:

هي عملية تهدف إلى استخراج المعلومات المهمة من البيانات في قواعد البيانات الكبيرة. يتبنى مفهوم استكشاف المعارف الطرق المطورة في تعلّم الآلة (MACHINE LEARNING)، والتنقيب عن البيانات (DATA MINING) التي توفر تقنيات لتحليل البيانات وكذلك الطرق المطورة في التعلم الإحصائي (STATISTICAL LEARNING). [5]

1. وسائل التنقيب في البيانات:

يعتبر التنقيب في المعطيات (DATA MINING) نواة استكشاف المعرفة في البيانات المهيكلة، وأثبت وجوده كأحد الحلول الناجحة لتحليل كميات ضخمة من البيانات. هناك عدة وسائل مختلفة من أجل التنقيب في البيانات، واختيار الوسيلة المناسبة يعتمد على طبيعة البيانات تحت الدراسة وعلى حجمها، بعض من هذه الوسائل هي:

- التفكير واستخلاص النتائج والقوانين من أمثلة حية (CASE–BASED REASONING): عملية حل المشاكل الجديدة القائمة على حلول لمشاكل مماثلة سابقة.
- الكشف عن قانون (RULE DISCOVERY): البحث عن منوال معين أو علاقة معينة في جزئية كبيرة من البيانات.
- معالجة الإشارات (SIGNAL PROCESSING): ايجاد الظواهر المتشابه مع بعضها البعض.
- الشبكات العصبية (NEURAL NETS): تطوير نماذج قابلة لتنبؤ النتائج، وهذه النماذج تم تطويرها بناءً على أسس تم استنباطها من عقل الإنسان، مثل تمييز الوجوه وتمييز الصوت. [6]

2. التقانات المستخدمة في تنقيب البيانات:

صممت أدوات التنقيب في البيانات لتضع الفرضيات وتختبرها وتستنتج منها معلومات جديدة، ويمكن القول إنه في حال معرفة المطلوب تماماً، يمكن استخدام لغات الاستفسار مثلSQL ، أما إذا كان ما يريده المستخدم غير واضح، ولديه فرضيات أو معايير، فلا بد من اللجوء إلى تقانات تنقيب في البيانات وثمة طيف واسع من التقانات المستخدمة في تنقيب البيانات يتميز كل منها بخصائص فريدة لا توفرها التقانات الأخرى ومن أهم هذه التقانات:

- أدوات الاستعلام (QUERY TOOLS) ومنها لغة SQL وغيرها. وهي الخطوة الأولى في عملية التنقيب في البيانات التي تستخدم أدوات استعلام تقليدية لمعرفة بنية المعطيات الأساسية بتطبيق خوارزميات تحليل معقدة.
- التقنيات الإحصائية (STATISTICAL TECHNIQUES) وهي طريقة مفيدة في استخلاص عينات من مجموعة البيانات، تطبق في بداية عملية التنقيب عن المعطيات.
- الإظهار (VISUALIZATION) تعطي فكرة جيدة حول كفاية البيانات والعينات التي يمكن استخلاصها منها.
- أشجار القرار (DECISION TREES) وهي تقانة مفيدة في تصنيف البيانات وفقاً لمجموعة من الخواص المتدرجة التي تشكل شجرة القرار، وتستخدم عادة أشجار القرار في الابحاث العملية، وتحديدا في تحليل القرار، للمساعدة في تحديد الاستراتيجية الأكثر احتمالا للوصول إلى الهدف.

- القواعد المترابطة (ASSOCIATION RULES) وسيلة لاكتشاف العلاقات بين المتغيرات مثيرة للاهتمام في قواعد البيانات الكبيرة. [7]

3. مراحل استكشاف المعرفة:

تتطلب عملية استكشاف المعرفة ست مراحل:

- اختيار البيانات (DATA SELECTION): ويجري في هذه المرحلة تجميع المعلومات المراد معالجتها من مخازن البيانات المختلفة ووضعها في مخزن للبيانات.
- التصفية (CLEANING): وتتضمن إلغاء التسجيلات المتكررة، وتصحيح أخطاء كتابة سلاسل المحارف، وإضافة المعلومات الناقصة وغيرها.
- الإغناء (ENRICHMENT).
- الترميز (CODING): إذ يجري استخدام ترميز وتصنيف موحد للبيانات ذات الدلالة المشتركة والمستوردة من بنوك المعطيات المعنية.
- التنقيب عن البيانات (DATA MINING).
- بناء التقارير (REPORTING). [8]

منهجية البحث:

إن تحقيق الهدف النهائي للويب الدلالي أمر معقد جداً ويتطلب الكثير من العمل بسبب الحجم الكبير للبيانات وترابطها بشكل مباشر أو غير مباشر. لكن في مثالنا البسيط اتبعنا المنهج الوصفي التحليلي لتوصيف الكيانات الموجودة لدينا وتحليل العلاقات المباشرة وغير المباشرة التي بينها لنصل إلى قاعدة بيانات متوافرة على الشبكة بهرمية منطقية ومناسبة للواقع. في البداية عملنا على توصيف هيكلية الجامعة من إدارة إلى أقسام إلى كليات، ومن ثم توصيف الإداريين، المدرسين، المواد والطلاب وتوصيف العلاقات بينهم وبين الجامعة، كربط المواد بالقسم الخاص بها وربط مدرسي هذه المواد بها وبالتالي بالأقسام التابعة لها. الآن يمكننا باستفسار بسيط الحصول على أسماء المدرسين لمادة ما، أو التابعين لقسم معين. كذلك يمكن الحصول على أسماء الطلاب الناجحين في مادة ما، أو اسم الطالب الحاصل على أعلى معدل وغيرها من الاستفسارات العديدة التي يمكن تطبيقها ببساطة وتوفير الجهد الكبير والكثير من الوقت للحصول على هذه النتائج من خلال البحث التقليدي في محركات البحث.

1. الحزمة البرمجية DotNetRDF:

قامت مجموعة من المطورين بتطوير مشروع DOTNETRDF، هدف هذا المشروع هو إنشاء مكتبة NET. مفتوحة المصدر تستخدم آخر إطار عمل من NET. لتزود المبرمجين بـ API قوية وسهلة الاستخدام ومناسبة لـRDF, SPARQL وللويب الدلالي بشكل عام، مما يمكن من توفير طريقة فعالة للعمل مع تقنيات الويب الدلالي في الـNET.

2. الصفوف الأساسية في الحزمة البرمجية DotNetRDF:

- **IGraph**

يمكن اعتبار مستند الـRDF أنه يشكل بيان رياضي، كل البيانات في الحزمة البرمجية DOTNETRDF تحقق الصف IGRAPH

- **INode**

كل بيان مكون من مجموعة عقد، ويوجد أنواع لهذه العقد: IBLANKNODE التي تعبر عن عقدة فارغة، ILITERALNODE وهي عقدة نصية يمكن أن تتعامل مع عدة لغات، IURINODE وهي عقدة من النمط URI.

3. الثلاثيات Triples:

الثلاثيات هي الوحدة الأساسية في بيانات RDF، فالعقد لا تملك أي معنى بدون استخدامها في ثلاثيات لتعطي معرفة، تنظم الثلاثيات في بنية (SUBJECT, PREDICATE AND OBJECT) والتي تمثل مكونات الجملة فاعل وفعل ومفعول به، وذلك يعني أن الفاعل (SUBJECT) يرتبط بالمفعول به (OBJECT) بواسطة الفعل والذي يمثل العلاقة بينهما (PREDICATE)، ويمثل كل منها عقدة.

4. مخازن الثلاثيات (Triple Store) في الحزمة DotNetRDF:

تستخدم مخازن الثلاثيات من أجل تمثيل مجموعة من الرسوم البيانية (GRAPHS) والسماح لنا بالعمل مع كميات كبيرة جداً من البيانات بشكل أسهل، وهي تعتمد على واجهة (ITRIPLESTORE) التي تعرف الخواص التالية لمخازن الثلاثيات:

- **Graphs**

تعيد مجموعة من الـGRAPHS في TRIPLE STORE قابلة للتجول عليها، حيث تكون هذه الـGRAPHS محملة في الذاكرة من أجل TRIPLE STORE معين، وليس بالضرورة أن نحمل كل الثلاثيات الموجودة في الـTRIPLE STORE.

- **IsEmpty**

تعيد نتيجة منطقية تبين هل مخزن الثلاثيات فارغ أم لا.

- **Triples**

تعيد مجموعة من الثلاثيات من كل الـGRAPHS الموجودة حالياً في الذاكرة والمسترجعة من TRIPLE STORE ما، مما يعني أنه ليس بالضرورة أن نسترجع كل الـ GRAPHS الموجودة في الـTRIPLE STORE.

5. الصفان الأساسيان للوصول إلى النتائج عند استخدام الـSPARQL:

هما SPARQLRESULTSET و SPARQLRESUL ، يقدم هذان الصفان RESULT SET و RESULT على التتالي، عند تنفيذ أي نوع استعلام SPARQL بواسطة أي طريقة سنحصل على SPARQLRESULTSET أو على IGRAPH كنمط معاد من الاستعلام.

يستخدم الصف SPARQLRESULTSET لتقديم نتائج الاستعلام عن طريق التعليمتين SELECT وASK ، إما أن تحوي RESULTSET جدول من عناصر نوعها SPARQLRESULT وذلك في حالة الاستعلام بواسطة SELECT، أو تحمل قيمة بوليانية في حالة الاستعلام عن طريق التعليمة ASK.

6. قاعدة المعرفة DBPEDIA:

تغطي طيف واسع من المعارف البشرية، سامحةً للمستخدمين بإيجاد المعلومات من مصدر واحد عوضاً عن تتبع نتائج محركات البحث، ومتابعة مئات الصفحات المنسوخة غالباً بحثاً عن المعلومات الجديدة. مثلاً لإيجاد عدد طلاب جامعة دمشق الحالي يوجد خيارين إما البحث بواسطة أحد محركات البحث المشهورة (GOOGLE, YAHOO,...) عن ذلك وإيجاد الصفحات التي تحوي هذه المعلومة واختيار أحدث تاريخ نشر لمقالة تتضمن ذلك، أما الخيار الثاني فهو الاستعلام

عن ذلك العدد من قاعدة المعرفة وبالتالي الحصول على أحدث وآخر تعديل، ويظهر جلياً أن خيار الاستعلام هو الأفضل والأسرع والأدق.

انطولوجيا البحث:

إن الفكرة الأساسية للبحث هي تزويد خدمات متعلقة بالتعليم العالي الحكومي السوري، مثل معلومات عن الجامعات الحكومية السورية كاسم الجامعة (الافتراضية، تشرين، حلب، دمشق، والبعث)، موقع الجامعة، عنوان موقعها على شبكة الويب، عدد الطلاب فيها، وملخص عن الجامعة، مما يساعد العملاء الأذكياء على إيجاد تلك الخدمات ديناميكياً، يمكن للعملاء بعدها أن تقترح الجامعة الأفضل تبعاً لتفضيلات المستخدم. بنية الويب الدلالي هنا ستعتمد على بضع أنطولوجيات أساسية، مثلاً الأنطولوجيا الخاصة بالجامعة الافتراضية السورية التي قمنا ببنائها في الموقع، وكذلك الأنطولوجيا الخاصة بالجامعات الأخرى الموجودة أصلاً في DBPEDIA. بالتالي يمكننا بناء مشروع يهدف إلى توفير محتوى مهيكل من المعلومات على شبكة الانترنت كما هو الحال في ال DBPEDIA تقدمه وزارة التعليم العالي سامحةً للمستخدمين بإيجاد المعلومات عن كافة الجامعات من مصدر واحد باستعلام بسيط عوضاً عن تتبع نتائج محركات البحث، ومتابعة مئات الصفحات المنسوخة غالباً بحثاً عن المعلومات الجديدة.

سنوضح جزء من الأنطولوجيا الخاصة بالجامعة الافتراضية السورية التي صممت ضمن ملفات RDF:

في البداية قمنا بتوصيف هيكلية الجامعة الافتراضية ضمن ملف RDFSCHEMA حيث عرفنا صف الجامعة وعرفنا خصائصه مثل: اسم الجامعة، البلد، المدينة، موقعها على الانترنت وغيرها كما يلي:

```
<-- تعريف صف الجامعة وخصائصه --!>
 <rdfs:Class                                    rdf:ID="University"/>
<rdf:Property                                  rdf:ID="UniversityName">
   <rdfs:domain                                 rdf:resource="#University"/>
   <rdfs:range          rdf:resource="http://www.w3.org/2000/01/rdf-schema#Literal"/>
                                                </rdf:Property>
<rdf:Property                                   rdf:ID="founded">
   <rdfs:domain                                 rdf:resource="#University"/>
<rdfs:range
                rdf:resource="http://www.w3.org/2000/03/example/classes#Number"/>
                                                </rdf:Property>
<rdf:Property                                   rdf:ID="country">
   <rdfs:domain                                 rdf:resource="#University"/>
   <rdfs:range          rdf:resource="http://www.w3.org/2000/01/rdf-schema#Literal"/>
                                                </rdf:Property>
<rdf:Property                                   rdf:ID="city">
   <rdfs:domain                                 rdf:resource="#University"/>
   <rdfs:range          rdf:resource="http://www.w3.org/2000/01/rdf-schema#Literal"/>
                                                </rdf:Property>
<rdf:Property                                  rdf:ID="AcademicSupervisor">
   <rdfs:domain                                 rdf:resource="#University"/>
   <rdfs:range          rdf:resource="http://www.w3.org/2000/01/rdf-schema#Literal"/>
                                                </rdf:Property>
<rdf:Property                                  rdf:ID="President">
   <rdfs:domain                                 rdf:resource="#University"/>
   <rdfs:range          rdf:resource="http://www.w3.org/2000/01/rdf-schema#Literal"/>
```

```
                                                    </rdf:Property>
<rdf:Property                                       rdf:ID="website">
  <rdfs:domain                                      rdf:resource="#University"/>
  <rdfs:range          rdf:resource="http://www.w3.org/2000/01/rdf-schema#Literal"/>
                                                    </rdf:Property>
```

ثم عرفنا صفوف الكلية، القسم والمواد وخصائص كل منها كما يلي:

```
<-- تعريف صف الكلية --!>
      <rdfs:Class                                   rdf:ID="Faculty"/>
 <rdf:Property                                      rdf:ID="FacultyName">
  <rdfs:domain                                      rdf:resource="#Faculty"/>
  <rdfs:range          rdf:resource="http://www.w3.org/2000/01/rdf-schema#Literal"/>
 </rdf:Property>


<-- تعريف صف القسم--!>
      <rdfs:Class                                   rdf:ID="Department"/>
 <rdf:Property                                      rdf:ID="DepartmentName">
  <rdfs:domain                                      rdf:resource="#Department"/>
  <rdfs:range          rdf:resource="http://www.w3.org/2000/01/rdf-schema#Literal"/>
 </rdf:Property>

<-- تعريف صف المنهاج--!>
      <rdfs:Class                                   rdf:ID="Course"/>
 <rdf:Property                                      rdf:ID="CourseName">
  <rdfs:domain                                      rdf:resource="#Course"/>
  <rdfs:range          rdf:resource="http://www.w3.org/2000/01/rdf-schema#Literal"/>
 </rdf:Property>
```

بعد ذلك قمنا بإعطاء قيم لهذه الخصائص كما يلي:

```
<-- إعطاء خصائص صف الجامعة قيم --!>
     <SVU:University                    rdf:about="http://www.svuonline.org">
   <SVU:UniversityName>The Syrian Virtual University</SVU:UniversityName>
                  <SVU:founded>2002</SVU:founded>
                  <SVU:country>Syria</SVU:country>
                  <SVU:city>Damascus</SVU:city>
                  <SVU:website>http://www.svuonline.org</SVU:website>
                  <SVU:President> Dr. Riad Daoudi </SVU:President>
                  <SVU:AcademicSupervisor>Dr. Bassel Al-
Khateb</SVU:AcademicSupervisor>
<-- إعطاء خصائص صف الكلية قيم --!>
        <SVU:faculties>
     <rdf:bag>
      <rdf:li>
       <SVU:Faculty                  rdf:about="#IT"                  />
      </rdf:li>
       <rdf:li>
        ISE
```

```
        </rdf:li>
        <rdf:li>
            HND
        </rdf:li>
        <rdf:li>
    </rdf:bag>
     </SVU:faculties>
          </SVU:University>
```

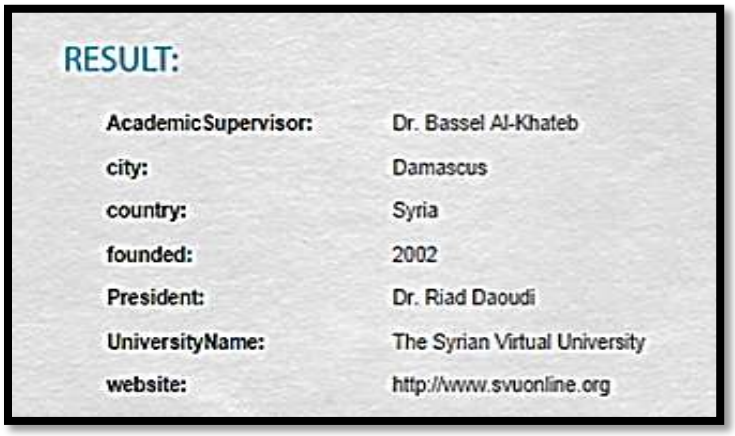
النتائج والمناقشة:

بعد بناء الأنطولوجيات وتمثيل الثلاثيات بمعايير RDF AND OWL، تبدأ مرحلة الحصول على معلومات بواسطة لغة الاستعلام SPARQL (SIMPLE PROTOCOL AND RESOURCE DESCRIPTION FRAMEWORK QUERY LANGUAGE). حيث تدعم SPARQL التجميع والاستعلامات الجزئية وغيرها من الأدوات التي تسهل عملية الحصول على المعرفة.

بعض استعلامات SPARQL المستخدمة في التطبيق ونتائجها:

- الاستعلام عن معلومات عامة عن الجامعة الافتراضية السورية ونتائج هذا الاستعلام كموقع الجامعة في دمشق، موقعها الافتراضي، تاريخ إنشائها، اسم رئيس الجامعة و مدير البرنامج، وهذه المعلومات تم الاستعلام عنها من ملفات ال RDF التي قمنا بتصميمها عن الجامعة الافتراضية والتي ممكن أن نضع فيها المعلومات التي نريدها وتصبح كبنية معلومات مهيكلة على الانترنت.

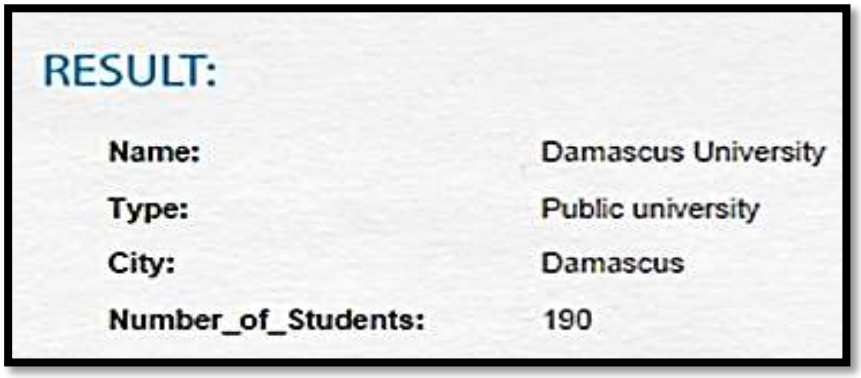
```
SELECT ?UNIVERSITYNAME ?FOUNDED ?COUNTRY ?CITY ?WEBSITE ?PRESIDENT
?ACADEMICSUPERVISOR
                WHERE {
    <HTTP://WWW.SVUONLINE.ORG>
<HTTP://ME.SOMEE.COM/SUNISCHEMA.RDF#UNIVERSITYNAME> ?UNIVERSITYNAME.
    <HTTP://WWW.SVUONLINE.ORG>
<HTTP://ME.SOMEE.COM/SUNISCHEMA.RDF#FOUNDED> ?FOUNDED.
    <HTTP://WWW.SVUONLINE.ORG>
<HTTP://ME.SOMEE.COM/SUNISCHEMA.RDF#COUNTRY> ?COUNTRY.
     <HTTP://WWW.SVUONLINE.ORG>  <HTTP://ME.SOMEE.COM/SUNISCHEMA.RDF#CITY>
?CITY.
    <HTTP://WWW.SVUONLINE.ORG>
<HTTP://ME.SOMEE.COM/SUNISCHEMA.RDF#WEBSITE> ?WEBSITE.
    <HTTP://WWW.SVUONLINE.ORG>
<HTTP://ME.SOMEE.COM/SUNISCHEMA.RDF#PRESIDENT> ?PRESIDENT.
    <HTTP://WWW.SVUONLINE.ORG><HTTP://ME.SOMEE.COM/SUNISCHEMA.RDF#ACADEM
ICSUPERVISOR> ?ACADEMICSUPERVISOR }
```



الشكل (2) يبين نتائج الاستعلام السابق كما تم اظهارها في تطبيق الويب المصمم

- استعلامات عن جامعة دمشق مأخوذة من ال dbpedia، مثل موقع الجامعة ونوعها وعدد الطلاب فيها ، وعنوان جامعة دمشق على الشبكة العنكبوتية مع ملخص عنها:

```
SELECT DISTINCT ?Name ?Type ?City ?Number_of_Students
WHERE {
        univ:" + university + @" rdfs:label ?Name.
        FILTER langMatches(lang(?Name), 'EN').
        OPTIONAL {
            univ:" + university + @" dbpedia-owl:type ?typeResource.
            ?typeResource rdfs:label ?Type.
            FILTER langMatches(lang(?Type), 'EN').
        }
        OPTIONAL {
            univ:" + university + @" dbpedia-owl:city ?cityResource.
            ?cityResource rdfs:label ?City.
            FILTER langMatches(lang(?City), 'EN').
        }
        OPTIONAL {
            univ:" + university + @" dbpedia-owl:numberOfStudents ?Number_of_Students
        }
    }
    LIMIT 1
```



الشكل (3) يبين نتائج الاستعلام السابق كما تم اظهارها في تطبيق الويب المصمم

نلاحظ سهولة الاستعلام عن بيانات الجامعات بدل من البحث عنها في محركات البحث وضياع الوقت والجهد بين صفحات الويب التي قد تقدم لنا بعض المعلومات التي نريد.

باستخدام تقنيات الويب الدلالي نستطيع استرجاع كل المعلومات المتاحة، خاصةً عند استخدام قواعد الاستدلال والتي تساعد في استرجاع بيانات ضمنية أي غير مصرح عنها بشكل مباشر، وكمثال عليها ليكن لدينا البيانات التالية:

PERSON MAN RDFS : SUBCLASSOF

MAN BASSAM RDF:TYPE

سيتم استخلاص العلاقة التالية والتي لم يصرح عنها مباشرة:

RDFS : SUBCLASSOF PERSON BASSAM

وتظهر أهمية الاستدلال ببعض الاستعلامات التي لا تعيد أي نتيجة ولكن بعد تطبيق قواعد الاستدلال يعيد الاستعلام العديد من النتائج.

الاستنتاجات والتوصيات:

توصلت الدراسة إلى النتائج الآتية:

1 -نستطيع أن نقول إن الويب الدلالي هو مفهوم فعال جداً لإدارة المعرفة واسترجاع البيانات.

2 -يمكن أن نستنتج من خلال تصميم التطبيق أن حلول الويب الدلالي توفر قدراً كبيراً من المرونة لاعتمادها من أجل تصنيف البيانات والتعديل عليها.

3 -إمكانية نشر أو تبادل البيانات على شكل ملفات RDF من خلال وضعها على شبكة الإنترنت أو من خلال توفير واجهة تقبل استفسارات SPARQL.

4 -يتطلب توظيف الويب الدلالي في بيئة المعرفة الاستعانة بالأنطولوجيات باعتبارها طريقة لتمثيل المفاهيم والربط بينها بعلاقات ذات معنى.

5 -استخدام قواعد الاستدلال Inference Rules تشكل إضافة مهمة لاسترجاع البيانات المطلوبة.

6 -تتيح لغة SPARQL الاستفهامية للتطبيقات البحثَ عن معلومات محدَّدة ضمن بيانات RDF مع إمكانية تطبيق بعض قواعد الاستدلال.

وفي ضوء تلك النتائج، توصي الدراسة بالآتي:

1 إجراء المزيد من الدراسات حول كيفية تطوير تطبيقات الويب الدلالي بما يناسب الآمال المتوقعة من الأسس النظرية الموافقة.

2 محاولة تقييم أداء الويب الدلالي مقابل الويب الحالي من أجل إبراز الهوية التكنولوجية للويب الدلالي.

3 بناء مشروع يهدف إلى توفير محتوى مهيكل من المعلومات على شبكة الانترنت كما هو الحال في DBPEDIA تقدمه وزارة التعليم العالي سامحةً للمستخدمين بإيجاد المعلومات عن كافة الجامعات من مصدر واحد عوضاً عن تتبع نتائج محركات البحث، ومتابعة مئات الصفحات المنسوخة غالباً بحثاً عن المعلومات الجديدة.

المراجع:

1- HEFLIN, J. D. *Towards the Semantic Web: Knowledge Representation in a Dynamic Distributed Environment*. University of Maryland, USA, 2001, 236.

2- WALTON, C. *Agency and the Semantic Web*. Oxford University Press, USA, 2007, 249.

3- Hawke, S; Herman, I; Archer, P; PRUD'HOMMEAW, E; Herman, I. *Semantic Web Activity*. W3C, 2013, 30 Sept, 2013. < http://www.w3.org/2001/sw/>

4- HITZLER, P; KRöTZSCH, M; PARSIA, B, PATEL-SCHNEIDER, P.F; RUDOLPH, S. *OWL 2 Web Ontology Language Primer*. World Wide Web Consortium, 2013, 415.

5- DžEROSKI, S; LAVRAC, N. *Relational Data Mining*. Springer, Berlin, 2010, 398.

6- LIAO, S; CHU, P; HSIAO, P. *Data mining techniques and applications–A decade review from 2000 to 2011*. Elsevier Ltd, 2012, 311.

7- AGGARWAL, G; CHARU, C; XIANG, C; ZHAI, M. *Mining text data*. Springer-Verlag, New York, 2012, 524.

8- PRUD'HOMMEAW, E; SEABORNE, A. *SPARQL query language for RDF*. W3C recommendation, 2008, 20 Oct, 2013. < http://www.w3.org/TR/rdf-sparql-query/>

9- ANGLES, R. GUTIERREZ, C. *Subqueries in SPARQL*. AMW, Chile, 2011, 250.

10- Swick, R; Schreiber, G; Wood, D. *Semantic Web Best Practices and Deployment Working Group*. W3C, 2006, 12 Feb, 2013. <http://www.w3.org/2001/sw/BestPractices>

11- Lehmann, J; Isele, R. *DBpedia – A Large-scale*. 10 Jan, 2014. <http://wiki.dbpedia.org/>