14/ محركات البحث الدلالي والأنطولوجيات :

إن ورود المعلومات على الإنترنت يزداد بشكل كبير، فقد أصبح الإنترنت مكانًا للتعبير عن الأفكار، سرد القصص، إنشاء المدونات ومشاركة الفيديوهات والصور والملفات الصوتية وما إلى ذلك. وهو ما جعل كمّ المعلومات المتوفرة للفرد الواحد أكبر بكثير مما يُمكن له أن يستفيد منه. تعريض العقل البشري لهذا الكم الهائل من المعلومات من شأنه أن يتسبب فيما يُمكن أن نصفه بوصف «الضياع في فضاء المعلومات»، وذلك راجع إلى بقاء المعلومات المفيدة بعيدة المنال بسبب تراكم الكثير من المعلومات غير المفيدة وغير المرتبطة بالموضوع المراد البحث عنه من قبل المستخدم. لحسن الحظ، مثلما يزداد ورود المعلومات، تزداد مقدرات معالجة المعلومات اوتوماتيكياً، لذا يوجد إمكانيات كبيرة للاستفادة من مقدرات الأتمتة هذه بهدف استخراج المعلومات والخدمات من فيضان الويب والمرتبطة بالمستخدم، وتوصيلها إليه عن طريق واجهة مستخدم معيارية .(Standardized User Interface) إن أهمية الحصول على المعلومات بهذه الطريقة التكيفية يزداد بازدياد كتلة المعلومات المتوفرة على الإنترنت.

تعتبر شبكة الويب أغنى المصادر المعلوماتية بما تحويه من مستندات ومعلومات ومصادر منوعة يمكن الوصول إليها عن طريق محركات البحث التقليدية. غير أن تنظيم هذه المعلومات والمستندات 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 ذات الدلالات والمعاني اللفظية»، أو ما يطلق عليه بالإنجليزية مصطلح (Semantic Web)، والتي هي امتداد للويب الحالية ولكن تختلف عنها بأنها تتفهم مدلولات الألفاظ والمعاني البشرية. يهدف البحث في الويب الدلالي إلى الإجابة عن أمرين أساسيين هما:

أولاً: للإجابة عن السؤال الأول ظهر مصطلح الأنطولوجيا الذي له أصله في علم الفلسفة، ففي الفلسفة يعني دراسة الكيانات والعلاقات بينها. واسقط الباحثون في مجال الويب الدلالي هذا المصطلح في حل مشاكلهم فاستخدموه للتعبير عن تمثيل رسمي للمعرفة على أنها مجموعة من المفاهيم في مجال محدد، والعلاقات بين هذه المفاهيم، ويمكن أن تتضمن الأنطولوجية وصفاً للصفوف (Classes) والكائنات (Objects) وخصائصها (Properties)والعلاقات بينها. ومن اللغات التي تستخدم لوصف الأنطولوجيات ,RDF) .RDF Schema, OWL).

ثانياً: للإجابة عن السؤال الثاني نعرض فيما يلي مثالاً بسيطاً لتطبيق للويب الدلالي .بمعنى أنه عند التواصل بين حاسوب وآخر النحو جدا مهم. إذا طلبت صفحة إنترنت فإنك تأمر السيرفر بتحميلها. ويبحث عنها ويظهرها في نسخة .html ولكن ما هو مكتوب على تلك الصفحات يبقى غير مفهوم له! نفس الشيء يحصل عند البحث على الإنترنت. المحرك

يبحث في مليارات الصفحات للكلمات التي أردت البحث عنها وبعدها يظهرها عبر صفحات في لائحة فيها الكلمة التي بحثت عنها ولكن المضمون وما أردت البحث عنه بالتحديد غير معروف أو مفهوم للمحرك. لهذا يمكننا القول بأن محرك البحث والحاسب مثل الببغاء يقولون ما نعلمهم إياه ولكن لا يفهمونه. الويب الحالي يمكن تسميته ويب الملفات. الملفات مرتبطة بشكل ذكي يبعضها البعض ولكن ما يوجد بداخلها يظل غامضاً.وبما أن المستخدم يبحث عن مضمون تلك المفات مثل التي تتحدث عن ناس، حوادث، شركات، بلدان، رياضة، طعام، تطور، علوم... مواضيع عديدة تتحدث عن أفراد... أو في كلمة واحدة مكونات.entity -

إذاً فالذكاء والتطوير هنا أن نُعرف الحاسب والويب عما يوجد داخل تلك الملفات !! هذا هو سيمنتاك ويب أو الويب الدلالي.إذا الويب الدلالي هو عكس الويب الحالي. هو ويب مع كيانات.

15 / التنقيب في البيانات واستخراج المعارف من المصادر :

/ تنقيب البيانات(Data mining) :

يشير معنى تنقيب البيانات Data mining إلى عملية استخراج الأنماط أو الأفكار المفيدة من كميات كبيرة من البيانات. ويتضمن استخدام تقنيات وخوارزميات مختلفة لاكتشاف الأنماط والاتجاهات المخفية داخل البيانات. بهدف استخراج معلومات قيمة يمكن استخدامها لاتخاذ القرار، حيث يساعد المؤسسات على الكشف عن رؤى قيمة، وتحسين عمليات صنع القرار، وتحديد أنماط سلوك العملاء، واكتشاف الأنشطة الاحتيالية، وتحسين العمليات، وتحسين أداء الأعمال بشكل عام. و يتضمن استخراج البيانات عدة خطوات، منها:

/ إعداد البيانات:

تتضمن هذه الخطوة جمع البيانات وإعدادها للتحليل. وقد يشمل تنظيف البيانات وتكاملها وتحويلها واختيار المتغيرات أو الميزات ذات الصلة.

/ استكشاف البيانات :

يتم استكشاف البيانات للحصول على فهم أفضل لخصائصها وعلاقاتها وأنماطها المحتملة. وتُستخدم الإحصائيات الوصفية وتقنيات التصور وطرق تحليل البيانات الاستكشافية لهذا الغرض.

/ تحديد النمط :

تتضمن هذه الخطوة تطبيق خوارزميات وتقنيات استخراج البيانات لتحديد الأنماط أو العلاقات أو الهياكل داخل البيانات. وتشمل التقنيات الشائعة المستخدمة في استخراج البيانات التصنيف، والانحدار، والتجميع، واستخراج قواعد الارتباط، والكشف عن الشذوذ.

/ تقييم الأنماط :

بمجرد تحديد الأنماط، فإنها تحتاج إلى تقييم لأهميتها وفائدتها. ويتضمن ذلك تقييم جودة وموثوقية الأنماط المكتشفة وتحديد قيمتها المحتملة في معالجة المشكلة أو الهدف المحدد.

/ نشر المعرفة :

تتضمن الخطوة الأخيرة تطبيق الأنماط أو المعرفة المكتشفة على الاستخدام العملي. وقد يتضمن ذلك إجراء تنبؤات، أو إنشاء نماذج، أو تحسين العمليات، أو اتخاذ قرارات مستنيرة بناءً على الأفكار المكتسبة من عملية استخراج البيانات. و

خوارزميات تنقيب البيانات عديدة ، وتختلف هذه الخوارزميات في الأساليب والتقنيات المستخدمة، ومنها:

/ تجميع البيانات(Clustering) :

تهدف هذه الخوارزميات إلى تقسيم مجموعة البيانات إلى مجموعات فرعية (عناصر مشابهة) بناءً على الأنماط الموجودة في البيانات.

/ تصنيف البيانات(Classification) :

تهدف هذه الخوارزميات إلى بناء نماذج تستطيع تصنيف البيانات إلى فئات محددة مسبقًا. وتستخدم الخوارزميات التي تعتمد على التصنيف مجموعة من الأمثلة المصنفة لتعلم النمط ومن ثم تصنيف البيانات الجديدة.

/ التجميع على مستوى الجمهور(Community Detection) :

تستخدم هذه الخوارزميات لتحديد تجمعات أو مجموعات طبيعية داخل الشبكات الاجتماعية أو الشبكات المعقدة الأخرى. وتهدف هذه الخوارزميات إلى اكتشاف الهياكل الاجتماعية والمجموعات المرتبطة ببعضها البعض.

/ استنتاج القواعد والتحليل الاستنتاجي(Association Rule Mining) :

تستخدم هذه الخوارزميات لاستخلاص القواعد والعلاقات الكامنة في البيانات. وتهدف هذه الخوارزميات إلى العثور على قواعد الترابط بين العناصر في مجموعة البيانات.

/ تحليل الشبكات(Network Analysis) :

تستخدم هذه الخوارزميات لفهم وتحليل الشبكات المعقدة والعلاقات بين العناصر. وتهدف هذه الخوارزميات إلى تحديد العقد المهمة والمجتمعات والبنية الهرمية وأهمية العلاقات الاجتماعية.

من الأسباب الرئيسية التي تبرز أهمية تنقيب البيانات، ما يلي:

/ **اتخاذ القرارات الذكية :** يحفز تنقيب البيانات من اتخاذ القرارات الذكية والمستنيرة. وذلك من خلال استخدام تقنيات تحليل البيانات واستخلاص الأنماط، حيث يمكن للمنظمات فهم البيانات بشكل أفضل واستخلاص الإشارات والتصورات الهامة التي يمكن أن تؤثر في عمليات القرار.

/ **اكتشاف الأنماط والاتجاهات :** يمكن لتنقيب البيانات أن يساعد في اكتشاف الأنماط والاتجاهات الكامنة في البيانات. وذلك عن طريق تحليل البيانات الضخمة، يمكن تحديد العلاقات والترابطات بين المتغيرات المختلفة واكتشاف الاتجاهات المستقبلية المحتملة.

/ **تحسين الأداء التشغيلي :**يعزز تنقيب البيانات من الأداء التشغيلي للمنظمات. وتحديد الفجوات والتحسينات المحتملة في العمليات التشغيلية وتحسينها لتحقيق أفضل كفاءة وفاعلية عن طريق تحليل البيانات.

/ **فهم العملاء وتحسين التسويق :** يساعد تنقيب البيانات في فهم العملاء بشكل أفضل وتحسين جهود التسويق. وتخصيص استراتيجيات التسويق وتوجيهاتها بشكل أكثر فعالية لزيادة رضا العملاء وتحقيق المزيد من المبيعات من خلال تحليل سلوك العملاء وتفضيلاتهم ومعرفة احتياجاتهم.

/ **اكتشاف الاحتياجات الناشئة والفرص :** يساهم تنقيب البيانات في اكتشاف الاحتياجات الناشئة والفرص الجديدة في السوق. وتحديد الاتجاهات الصاعدة والفرص الناشئة واستغلالها للحصول على ميزة تنافسية وتحقيق نجاح أعلى في السوق.

هناك بعض الأهداف الشائعة لتنقيب البيانات، منها:

استكشاف البيانات وفهمها ، الكشف عن الأنماط والاتجاهات ،الاستنتاج والتنبؤ ،دعم اتخاذ القرار ،

توفير الدعم لاتخاذ القرارات،الاكتشاف والابتكار ،تحسين العمليات والكفاءة ،

/ أدوات تنقيب البيانات :

هناك العديد من الأدوات المستخدمة في تنقيب البيانات، والتي يمكن أن تختلف في القدرات والميزات التي تقدمها، ومنها:

/ منصةMonkeyLearn : هي منصة تعلم آلي ومعالجة لغوية طبيعية (NLP) تهدف إلى مساعدة الشركات في استخراج البيانات وتحليلها بشكل أوتوماتيكي. وتستخدم MonkeyLearnتقنيات التعلم الآلي لتحويل النصوص والبيانات غير المنظمة إلى بيانات منظمة ومفهومة.

/ منصة RapidMiner : هي منصة تحليل بيانات متكاملة وقوية تستخدم لتعدين البيانات وتحليلها، حيث تهدف RapidMiner إلى توفير بيئة تطوير مرئية وسهلة الاستخدام

للمحللين والعلماء في مجال البيانات لاستكشاف وتحليل البيانات واكتشاف الأنماط والتوجهات.

/ أداةOracle Data Mining :أداة Oracle Data Mining (ODM) هي إضافة متكاملة لقاعدة بياناتOracle Database ، وتعتبر جزءًا من Oracle Advanced Analytics.تم تصميم ODM لتوفير مجموعة من الوظائف والأدوات لتعدين البيانات وتحليلها داخل بيئة قاعدة البيانات.

أداةIBM SPSS Modeler : أداة IBM SPSS Modeler هي أداة تحليل بيانات وتعدين البيانات قوية وشاملة، والتي تم تطويرها بواسطة شركة .IBM وتسمح SPSS Modeler للمستخدمين بتحليل البيانات واستكشاف الأنماط وبناء نماذج تنبؤية بشكل سهل وفعال. وتتميز SPSS Modeler بواجهة رسومية سهلة الاستخدام وتعتمد على سحب وإفلات العناصر والعمليات المختلفة لبناء سلاسل تحليل البيانات.

/ برنامجWeka : هي برنامج حر ومفتوح المصدر لتحليل البيانات وتعدين البيانات. وتم تطويرها في جامعة وايكاتو في نيوزيلندا، وتوفر مجموعة واسعة من الخوارزميات والأدوات لتحليل البيانات وبناء نماذج تنبؤية. وتتميز Weka بواجهة مستخدم سهلة الاستخدام ومرئية، مما يجعلها مناسبة للمستخدمين ذوي الخبرة المختلفة في مجال التحليل البياني.

/ أداةKnime : أداة KNIME هي بيئة تحليل بيانات وتعدين بيانات قوية ومفتوحة المصدر، حيث تم تطويرها بواسطة شركةKNIME AG ، وتوفر مجموعة واسعة من الوظائف والأدوات لتحليل البيانات وبناء سلاسل التحليل. وتتميز KNIME بواجهة رسومية سهلة الاستخدام وقوية، وتعتمد على سحب وإفلات العناصر والعمليات المختلفة لبناء سلاسل التحليل.

/ أداةOrange : عبارة عن مجموعة أدوات متعددة الاستخدامات تُستخدم على نطاق واسع للأغراض التعليمية وتحليل البيانات والبحث. وهي متاحة لمنصات متعددة ولديها مجتمع نشط يساهم في تطويرها.

/لغةPython : البايثون هي لغة برمجة عالية المستوى ومفتوحة المصدر، ولها شعبية كبيرة في مجالات مختلفة بما في ذلك تحليل البيانات والعلوم البيانية. ويمكن استخدام Pythonلإنشاء برامج وتطبيقات متنوعة، ولديها مجموعة واسعة من المكتبات والأدوات المتاحة لتسهيل تحليل البيانات وتعدينها.

/ .أداةSAS Enterprise Miner : هي أداة تحليل بيانات قوية تقدمها شركة SAS Institute، حيث تم تصميمها خصيصًا لتعدين البيانات وتحليلها واستكشاف النمط والتنبؤ. وتستخدم SAS Enterprise Miner في مجالات متنوعة مثل الأعمال التجارية والصناعة والرعاية الصحية والتسويق والمالية والعلوم الاجتماعية.

/ لغةR : هي لغة برمجة وبيئة تطوير مفتوحة المصدر مصممة خصيصًا للتحليل الإحصائي والتصور البياني. وتعتبر R من أدوات التحليل الاحصائي الأكثر شعبية واستخدامًا في العديد من المجالات مثل العلوم الاجتماعية وعلوم البيئة والصحة والتسويق وعلوم البيانات.

ملاحظة مهمة : لفهم كل هذه البيانات و تطبيقها كباحث مستقبلا ، من لازم أخذ دورات تدريبية .

بالتوفيق