

مقارنة بين تطبيق نظم تعدين الآراء باللغة العربية وتطبيقها باللغة الانجليزية

عاصم سيداحمد على عثمان

عوض حاج علي

جامعة النيلين

مجلة كلية الدراسات العليا

الرقم الدولي الموحد: 1858-6228

المجلد: 15 ، 2020م

العدد: 11



كلية الدراسات العليا
جامعة النيلين

مقارنة بين تطبيق نظم تعدين الآراء باللغة العربية وتطبيقها باللغة الانجليزية

عاصم سيداحمد على عثمان، عوض حاج علي

asalageed@su.edu.sa

كلية علوم الحاسوب وتقانة المعلومات، جامعة النيلين، الخرطوم، السودان

المستخلص

هذه الورقة توضح مفهوم تعدين الآراء والادوات المستخدمة في تعدين الآراء، كما تناقش تطبيق نظام تعدين الآراء على خدمة الايداع النقدي لبنك فيصل الاسلامي السوداني باللغة العربية، وتطبيقه على خدمة سداد الرسوم الدراسية عبر البنك في جامعة النيلين باللغة الانجليزية كما تقوم بعمل مقارنة بين تطبيق انظمة تعدين الآراء باللغة العربية وتطبيقها باللغة الإنجليزية وتوضح المشاكل التي تواجه تطبيق مثل هذه الانظمة باعتبارها واحدة من الانظمة الحديثة والمجالات الحديثة في مجال علوم الحاسب الآلي. يمكن ان تطبق مثل هذه التقنية في عدد من المجالات منها التعليم، والتسويق، والصحة، والترفيه... الخ كما يمكن الاستفادة منها في جميع مناحي الحياة لأنها تساعد المؤسسات في اتخاذ قرارات مثلى، وهذه القرارات تبني على معلومات تم جمعها من المستفيدين بصورة عامة حيث تمت كتابتها بصورة تلقائية في شكل آراء.

مقدمة

يعتبر بمثابة تحدي كبير لأنها تقوم بمعالجة اللغة الطبيعية (natural language processing or text mining problem).

تعدين الآراء (Opinion Mining):

تعدين الآراء (Opinion Mining): هو علم ظهر مؤخراً ويتعامل اساساً مع استرجاع المعلومات واللغويات الحاسوبية والذي يعطى اهمية للرأي اكبر من الموضوع نفسه [1].

تعدين الآراء او تحليل المشاعر (Opinion Mining or Sentiment Analysis): يحلل ويركز على المعلومات الموضوعية في المصدر من خلال تطبيق معالجة اللغات الطبيعية، واللغويات الحاسوبية وتحليلات النص [2].

تعدين الرأي هو عملية الاستخراج التلقائي للمعرفة من رأي الآخرين حول بعض الموضوعات أو مشكلة معينة. فكرة تعدين الآراء وتحليل المشاعر هو "معالجة مجموعة من نتائج البحث لبند معين، وتوليد قائمة من سمات المنتج (جودة، ميزات الخ) وتجميع الرأي". ولكن مع مرور الوقت هناك تطبيقات أكثر إثارة للاهتمام والتطورات في هذا المجال والآن هدفها الرئيسي هو جعل الكمبيوتر قادراً على التعرف وتوليد العواطف مثل الإنسان [3].

النمو السريع لشبكة الإنترنت العالمية (www) جعلت المعلومة متاحة بصورة أفضل من السابق وبالتالي فإن الناس الآن وعلي نحو متزايد يأخذون معلوماتهم المطلوبة من بعضهم البعض وليس من الشركات أو وسائل الإعلام والهيئات السياسية وغيرها. فالإنترنت الآن أصبح وسط إجتماعي ذو شعبية كبيرة بحيث يغطي معظم أشكال المشاركات مثل التجارب والصور والتوصيات وغيرها، وللقيام بذلك أصبح الناس منخرطين بصورة كبيرة في الشبكات الإجتماعية المكونة من قوائم الأصدقاء. المدونين الذين يعلقون علي موضوع معين او عن طريق مستخدمين يكتبون بشكل تعاوني في موقع ويكيبيديا، وربما يبدي الناس آرائهم علي المشاركات وهذه الآراء يمكن ان تكون إيجابية أو سلبية أو موضع خلاف لهذه المشاركة.

إن إيجاد مصادر الآراء ومراقبتها علي الشبكة يمكن ان تكون مهمة شاقة لأن هناك عدد كبير من المصادر المختلفة وكل مصدر يحتوي علي حجم كبير من الآراء والمواقف ، وفي حالات عدة يمكن ان تكون الآراء مخبأة في المنتديات والمدونات وبالتالي يصعب علي القراء الحصول علي مصادر مفيدة ومتعلقة بالموضوع وقراءته وتلخيصه وتنظيمه بصورة صالحة للإستخدام، وهكذا فإننا نحتاج إلي أنظمة تقوم تلقائياً بإيجاد الآراء وتلخيصها. وتحليل ميول الناس (sentiment analysis) والذي يسمى أيضاً بتعدين الرأي (opinion mining) ظهر من اجل هذه الحاجة ولكن تطبيقها في الواقع

أدوات تعددين الآراء:

الأدوات التي يتم إستخدامها لمتابعة ومعرفة القطبية للآراء المتولدة من المستخدمين هي:

1. أداة مراجعة السير (Review Seer tool): يتم استخدام هذه الأداة لإكمال العمل الذي قامت به مواقع التجميع. يتم استخدام نهج المصنف Naive Bayes لجمع الآراء الإيجابية والسلبية لتعيين درجة استخراج ميزة الشروط [4].

2. النافورة الإلكترونية (Web Fountain): وتستخدم ابتداء من قاعدة محددة (bBNP) وهي عبارة عن نهج ارشادي لاستخراج ميزات المنتج [5].

3. أوبال الأحمر (Red Opal): هو الأداة التي تمكن المستخدمين من تحديد توجهات الرأي للمنتج بناء على خصائصها. وذلك بتعيين علامات لكل منتج على أساس الخصائص المستخرجة من استعراضات العملاء [6].

4. رأي المراقب (Opinion observer): هو نظام تعددين الرأي لتحليل ومقارنة الآراء حول الإنترنت باستخدام المحتويات التي ولدت بواسطة المستخدم. هذا النظام يظهر النتائج في شكل رسم بياني لرأي ميزة المنتج [5].

تعمل هذه الأدوات الآلية مع بعضها البعض وهناك العديد من أدوات الإنترنت مثل (تويتير وتوينز وغيرها ..) وهي متاحة لمراقبة الآراء في شبكة الإنترنت.

النتائج

- تطبيق مثل هذه الانظمة على المؤسسات سيساعد كثيرا في تقديم خدمات افضل وذلك بالاستفادة من الآراء او التعليقات التي يكتبها العملاء حول الخدمات التي تقدم لهم، وبالنسبة للعملاء تساعد مثل هذه الانظمة في اتخاذ قرار شراء المنتج او طلب الخدمة المعينة وذلك بالاطلاع على اراء المستخدمين السابقين.

- عند تطبيق النظام باللغة الانجليزية لاحظنا ان النظام قد اظهر نتائج فعالة او اكثر دقة من النتائج التي تم الحصول عليها عند تطبيق النظام باستخدام اللغة العربية وذلك بسبب مشاكل اللغة العربية الآتية:

- يمكن الحكم على ان اللغة العربية تعتبر من اصعب اللغات عند التعامل معها خاصة في مجال تكنولوجيا المعلومات وذلك بسبب تركيبة الجمل العربية حيث لا يمكن ان نحصر كل الكلمات التي يمكن ان تصنف سلبية او ايجابية وذلك يرجع مثلا (اذا كتبت كلمة بهمزة وبدون همزة فإن الحاسب الآلي سيواجه مشكلة لأنه يعتبر ذلك وجود فرق بين الكلمتين في حين ان المعنى واحد وبعض الاشخاص يكتبونها بطريقة والبعض الاخر يكتبها بطريقة اخرى).

- بعض الادوات المستخدمة للنفي في اللغة العربية في بعض الجمل لا تعني نفي بل تعني لكن مثل (الخدمة جميلة غير انها تاخذ وقتا طويلا) وفي جملة اخرى (الخدمة غير مفيدة) في الجملة الثانية تم استخدام كلمة غير كاداة نفي لكلمة مفيدة ولكن في الجملة الاولى لم تستخدم كاداة نفي وانما استخدمت بمعنى (لكن)، بالتالي يصعب تغذية الحاسب الالى بمثل هذه الكلمة كاداة للنفي لتغيير تصنيف الجملة من سلمي الى ايجابي او العكس.

- كثير من كلمات اللغة العربية لها اكثر من معنى بالتالي يصعب تحديد بعض الكلمات على انها ايجابية او سلبية وفي حالة عدم تحديدها نجد ان النظام سيقوم بتصنيفها الى محايد.

- يصعب نمذجة الكلمات العربية بنماذج رياضية مما يسهل على النظام فرزها الى رأي ايجابي او سلمي او محايد.

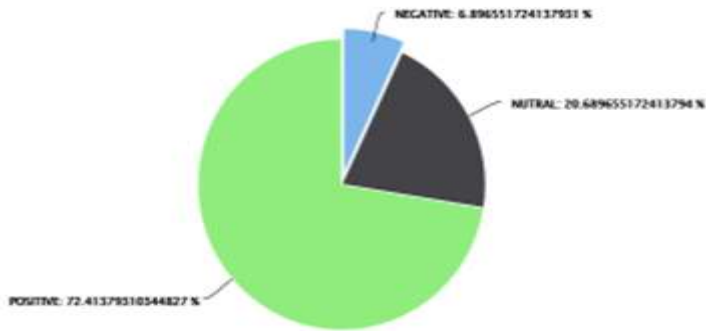
- استخدام تقييم النجوم ساعد كثيرا في تصنيف التعليق او الراي الى رأي ايجابي او سلمي او محايد كما يمكن ايضا استخدامه على حدا كتقرير منفصل يساعد على تقييم المنتج او الخدمة بالنسبة للعملاء الجدد.

- للناس اساليب مختلفة في الكتابة، وقد تحتوي نفس الجملة على رأي إيجابي ، فضلا عن رأي سلمي، لذلك فإنه من الصعب تحليل الجملة إيجابيا أو سلبيا في حالة تعددين الرأي على مستوى الجمل.

- استخدام تقنية الفرز بالجمل لها عيوب حيث ان هنالك عدد من الكلمات المركبة والتي لا يمكن فصلها من بعضها البعض مثل كلمة (Well come) على حسب النظام فإنه يتعامل مع كلمة

This Chart Created On 29 opinions

Comment Pie Chart



شكل رقم (1) يوضح تقرير pie chart لخدمة السداد عبر البنك حسب التعليقات او الآراء

- نسبة الآراء السلبية حول خدمة سداد الرسوم الدراسية عبر البنك هي (6.89%).
 - نسبة الآراء التي لم يتمكن النظام بتحليلها أي لم يتمكن من تصنيفها الى آراء ايجابية او سلبية هي (20.68%).
 - نسبة الآراء الايجابية حول خدمة سداد الرسوم الدراسية عبر البنك هي (72.41%).
- ولقياس كفاءة النظام تم اجراء تحليل للأوراق التي تم جمعها بطريقة يدوية حيث قام الباحث بتصنيف الآراء الى ثلاثة كما في النظام (ايجابية ، سلبية ، محايدة) وتم مراجعة كل ورقة وتحديد التصنيف الذي تتبع اليه هذه الورقة حيث وجد أن :

- عدد الاشخاص الذين كان رأيهم ايجابي عددهم (19) بالتالي فإن نسبة الآراء الايجابية حسب التحليل اليدوي هي (63.33%).
 - عدد الاشخاص الذين كان رأيهم سلبى عددهم (3) بالتالي فإن نسبة الآراء السلبية حسب التحليل اليدوي هي (10%).
 - عدد الاشخاص الذي تم اعتبار رأيهم محايد عددهم (8) بالتالي فإن نسبة الآراء المحايدة حسب التحليل اليدوي هي (26.66%).
- وبمقارنة النسب اعلاه مع النسب التي تم الحصول عليها من النظام نجد أن:

نوع الآراء	النسبة حسب النظام	النسبة حسب التحليل اليدوي
الآراء الايجابية	72.41 %	63.33 %

well لوحدتها وكلمة come بالتالي فإن النتيجة التي سنحصل عليها عن طريق النظام في بعض الاحيان لا تكون دقيقة.

- طرق المعجم الأساسي تستخدم لتعدين الآراء وهي وسيلة غير فعالة للتعامل مع كلمات النصوص التابعة. على سبيل المثال كلمة "صغير" يمكن أن تعبر عن رأى سلبى أو إيجابى وفقاً لميزات المنتج، للهاتف المحمول إذا قال العميل " ان حجم الهاتف صغير" هذه الجملة لا يمكن تصنيفها على أنها ذات رأى إيجابى أو سلبى.

- عند تطبيق النظام باللغة الانجليزية لوحظ كثرة وجود الاختصارات للدلالة على كلمات مما يصعب عملية الفرز الآلى.

- تجميع مرادفات الكلمات يعتبر مهمة صعبة حيث يجب تغذية النظام بجميع المترادفات حتى يتسنى له الحكم او تحديد الآراء الايجابية والسلبية.

مناقشة الدراسة:

تم تطبيق الدراسة على خدمتين مختلفتين مؤسستين ذات مجالات مختلفة ايضاً، حيث تم تطبيقها على شريحة من طلاب جامعة النيلين وذلك بأخذ آرائهم حول خدمة سداد الرسوم الدراسية عبر البنك، كما تم تطبيقها على شريحة من عملاء بنك فيصل الاسلامي السوداني وذلك بأخذ آرائهم حول خدمة الايداع النقدي عبر الصراف الآلى.

تطبيق نظام تعدين الآراء على خدمة سداد الرسوم الدراسية لطلاب جامعة النيلين عبر البنك باستخدام اللغة الانجليزية:

واحدة من الخدمات التي تقدمها ادارة تقانة المعلومات والاتصالات بجامعة النيلين هي خدمة سداد الرسوم الدراسية عبر البنك حيث يتم تحديد رسوم الطالب باستخدام نظام التسجيل ومن ثم يذهب الطالب الى البنك لسداد الرسوم الدراسية عبر البنك أي يتم السداد عبر البنك [7]، تم تطبيق النظام على هذه الخدمة حيث تم استهداف عدد (30) طالب من طلاب جامعة النيلين بصورة عشوائية لكتابة آرائهم حول الخدمة حيث طلب من كل واحد كتابة رايه حول الخدمة باستخدام اللغة الانجليزية في جملة لا تتجاوز (10) كلمات، ومن ثم تم إدخال هذه الاستمارات الى النظام.

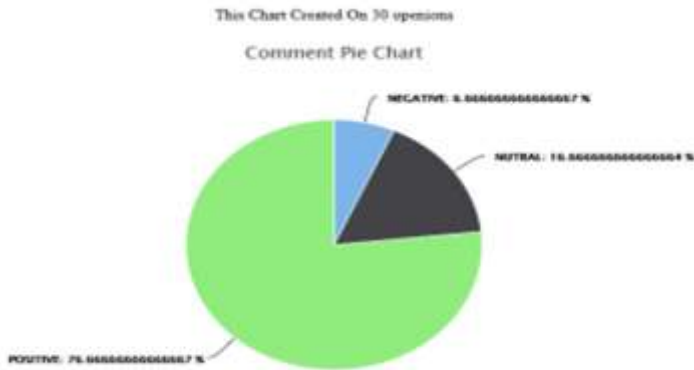
نتائج تطبيق النظام على خدمة سداد الرسوم الدراسية عبر البنك:

تطبيق نظام تعدين الآراء على خدمة الإيداع النقدي عبر الصراف الآلي لبنك فيصل الاسلامي باستخدام اللغة العربية:

يعتبر بنك فيصل الإسلامي السوداني هو البنك الأول الذي يقدم خدمة الإيداع عبر الصراف الآلي في السودان و تمكن هذه الخدمة عملاء البنك الكرام من إيداع المبالغ النقدية عبر صرافات البنك.[8].

تم تطبيق نظام تعدين الآراء على خدمة الإيداع النقدي عبر الصراف الآلي لبنك فيصل الاسلامي حيث تم استهداف عدد (30) شخصا من عملاء البنك من مختلف الجنسين حيث طلب من كل واحد كتابة رايه حول الخدمة باستخدام اللغة العربية (والدارجية السودانية اذا اراد) بجملة لا تتعدى (10) كلمات وتم ادخال الاستمارات على النظام.

نتائج تطبيق النظام على خدمة الإيداع النقدي عبر الصراف الآلي لبنك فيصل الاسلامي السوداني:



شكل رقم (2) يوضح تقرير pie chart لخدمة الإيداع النقدي عبر الصراف الآلي حسب الآراء

- نسبة الآراء السلبية حول خدمة الإيداع النقدي عبر الصراف الآلي هي (6.66%).
- نسبة الآراء التي لم يتمكن النظام بتحليلها أي لم يتمكن من تصنيفها إلى آراء ايجابية او سلبية هي (16.66%).
- نسبة الآراء الايجابية حول خدمة الإيداع النقدي عبر الصراف الآلي هي (76.66%).

الآراء السلبية	6.89%	10%
الآراء المحايدة	20.68	26.66

جدول يوضح نسب التحليل للآراء بالتطبيق على خدمة سداد الرسوم الدراسية عبر البنك

من النسب اعلاه نلاحظ ان هنالك فرق بين التحليل الذي تم اجرائه بصورة يدوية وبين التحليل الذي تم باستخدام النظام، وفي نظر الباحث ان النسبة يمكن ان تعتبر قليلة في حالة انها لم تتعدى نسبة 10%، ويرجع سبب الاختلاف في التحليلين لعدد من الاسباب هي:

1. تم استخدام اللغة الانجليزية عند التطبيق على هذه الخدمة وطلب من الطلاب كتابة آرائهم باللغة الانجليزية ولوحظ ان هنالك عدد من الطلاب الذين قاموا بكتابة كلمات بصورة املائية خاطئة بالتالي يصعب على النظام فرزها.
2. في الطريقة اليدوية تم تحليل الاوراق بحيث اذا غلبت الكلمات الايجابية تم تصنيف الراي الى راي ايجابي والعكس اذا غلب عدد الكلمات السلبية في التعليق على الكلمات الايجابية اعتبر الراي ايجابي.
3. بالنسبة للآراء المحايدة تم اعتبار الراي محايد عند تساوي عدد الكلمات الايجابية والسلبية في التعليق.
4. تجميع مرادفات الكلمات يعتبر مهمة صعبة حيث يجب تغذية النظام بجميع المترادفات حتى يتسنى له الحكم او تحديد الآراء الايجابية والسلبية، فمثلا اذا علق احدهم على الخدمة كالآتي: إذا علق العميل (A) " the Service quality is excellent" وعلق العميل (B) " quality of Service is very good" نجد أن الراين يصنفا آراء ايجابية ولكن تم استخدام كلمات مختلفة ولكنها تدل على نفس المعنى.

5. بالنسبة لتعليق العملاء بشكل حر يتم استخدام الاختصارات والكلمات القصيرة (على سبيل المثال استخدام كلمة cam تعبير عن الكاميرا وغيرها ..) وهذا النوع من اللغة بحاجة إلى الكثير من العمل لتعدين الرأي

ولقياس كفاءة النظام تم اجراء تحليل للأوراق التي تم جمعها بطريقة يدوية حيث قام الباحث بتصنيف الآراء الى ثلاثة كما في النظام (ايجابية ، سلبية ، محايدة) وتم مراجعة كل ورقة وتحديد التصنيف الذي تتبع اليه هذه الورقة حيث وجد أن :

- عدد الاشخاص الذين كان رأيهم ايجابي عددهم (15) بالتالي فإن نسبة الآراء الايجابية حسب التحليل اليدوي هي (50%).
 - عدد الاشخاص الذين كان رأيهم سلبى عددهم (7) بالتالي فإن نسبة الآراء السلبية حسب التحليل اليدوي هي (23.33%).
 - عدد الاشخاص الذي تم اعتبار رأيهم محايد عددهم (8) بالتالي فإن نسبة الآراء المحايدة حسب التحليل اليدوي هي (26.66%).
- وبمقارنة النسب اعلاه مع النسب التي تم الحصول عليها من النظام نجد أن:

نوع الآراء	النسبة حسب النظام	النسبة حسب التحليل اليدوي
الآراء الايجابية	76.66%	50%
الآراء السلبية	6.66%	23.33%
الآراء المحايدة	16.66%	26.66%

جدول يوضح نسب التحليل للآراء الخاصة بخدمة الايداع النقدي عبر الصراف الآلى

من النسب اعلاه نلاحظ ان هنالك فرق بين التحليل الذي تم اجرائه بصورة يدوية وبين التحليل الذي تم باستخدام النظام، ونلاحظ في هذه المرة ان النسب التي تم الحصول عليها من التحليلين متفاوتة بدرجة كبيرة اذ يتضح الفرق الشاسع بين التحليل الذي تم اجرائه بصورة يدوية وبين الذي تم اجرائه باستخدام النظام ، ويرجع سبب الاختلاف في التحليلين لعدد من الاسباب هي:

1. تم استخدام اللغة العربية عند التطبيق على هذه الخدمة وطلب من عملاء البنك كتابة آرائهم باللغة العربية حيث لوحظت عدد من المشاكل:

- يمكن الحكم على ان اللغة العربية تعتبر من اصعب اللغات عند التعامل معها خاصة في مجال تكنولوجيا المعلومات وذلك بسبب تركيبة الجمل العربية حيث لا يمكن ان نحصر كل الكلمات التي يمكن ان تصنف سلبية او ايجابية وذلك يرجع مثلا (اذا كتبت كلمة بهمزة وبدون همزة فإن الحاسب الآلى سيواجه مشكلة لأنه يعتبر ذلك وجود فرق بين الكلمتين في حين ان المعنى واحد وبعض الاشخاص يكتبونها بطريقة والبعض الاخر يكتبها بطريقة اخرى).
 - بعض الادوات المستخدمة للنفي في اللغة العربية في بعض الجمل لا تعني نفي بل تعني لكن مثل (الخدمة جميلة غير انها تاخذ وقتا طويلا) وفي جملة اخرى (الخدمة غير مفيدة) في الجملة الثانية تم استخدام كلمة غير كاداة نفي لكلمة مفيدة ولكن في الجملة الاولى لم تستخدم كاداة نفي وانما استخدمت بمعنى (لكن)، بالتالى يصعب تغذية الحاسب الآلى بمثل هذه الكلمة كاداة للنفي لتغيير تصنيف الجملة من سلبى الى ايجابى او العكس.
 - كثير من كلمات اللغة العربية لها اكثر من معنى بالتالى يصعب تحديد بعض الكلمات على انها ايجابية او سلبية وفي حالة عدم تحديدها نجد ان النظام سيقوم بتصنيفها الى محايد.
2. في الطريقة اليدوية تم تحليل الاوراق بحيث اذا غلبت الكلمات الايجابية تم تصنيف الراى الى راى ايجابى والعكس اذا غلب عدد الكلمات السلبية في التعليق على الكلمات الايجابية اعتبر الراى ايجابى.
 3. بالنسبة للآراء المحايدة تم اعتبار الراى محايد عند تساوى عدد الكلمات الايجابية والسلبية في التعليق.
 4. طرق المعجم الأساسى تستخدم لتعدين الآراء وهي وسيلة غير فعالة للتعامل مع كلمات النصوص التابعة. على سبيل المثال كلمة "صغير" يمكن أن تعبر عن رأى سلبى أو إيجابى وفقا لميزات المنتج، للهاتف المحمول إذا قال العميل " ان حجم الهاتف صغير" هذه الجملة لايمكن تصنيفها على أنها ذات رأى إيجابى أو سلبى.
 5. الجمع بين الآراء والسلوك للتحقق من صحة البيانات وتوفير مزيد من التحليل للبيانات من الآراء المعبرة.

8- خدمة الايداع النقدي عبر الصراف الآلي ، بنك فيصل الاسلامي
السوداني

المصادر والمراجع

www.fibsudan.com

- 1- Bing Liu . Exploring User Opinions in Recommender Systems. Proceeding of the second KDD workshop on Large Scale Recommender Systems and the Netflix Prize Competition, Aug 24, 2008, Las Vegas, Nevada, USA.
- 2- Alina Andreevskaia and Sabine Bergler. 2006. Mining WordNet for fuzzy sentiment: Sentiment tag extraction from WordNet glosses. In Proceedings of EACL-06, 11th Conference of the European Chapter of the Association for Computational Linguistics, Trento, IT. Forthcoming.
- 3- S Padmaja¹ and Prof. S Sameen Fatima², (2013), " Opinion Mining and Sentiment Analysis –An Assessment of Peoples' Belief: A Survey" in International Journal of Ad hoc, Sensor & Ubiquitous Computing (IJASUC) Vol.4, No.1
- 4- G.Vinodhini and RM. Chandrasekaran 2012. Sentiment analysis and Opinion Mining: A survey International Journal of advanced Research in Computer Science and Software Engineering vol. 2 Issue 6.
- 5- Yi and Niblack 2005. Sentiment Mining in Web Fountain" ,Proceedings of 21st international Conference on Data Engineering, pp. 1073-1083, Washington DC.
- 6- Christopher Scaffidi, Kevin Bierhoff, Eric Chang, Mikhael Felker, Herman Ng and Chun Jin 2007. Red Opal: product-feature scoring from reviews, Proceedings of 8th ACM Conference on Electronic Commerce, pp. 182-191, New York.

7- ادارة تقانة المعلومات والاتصالات ، جامعة النيلين

www.neelain.edu.sd