مجلة جامعة تشرين للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (40) العدد (6) 2018
Tishreen University Journal for Research and Scientific Studies - Engineering Sciences Series Vol. (40) No. (6) 2018

تطوير بيئة انترنت الأشياء (IOT) باستخدام نظام النشر والاشتراك

د. عبد الكريم السالم •
وائل حبيب••

(تاريخ الإيداع 23 / 9 / 2018. قُبِل للنشر في 26/ 12 / 2018)

□ ملخّص □

تستخدم حساسات انترنت الأشياء(internet of things) (IOT) نموذج النشر\ الاشتراك كوسيلة للاتصال مستفيدة من خصائصه المستقلة المتعلقة بالمكان والزمان والمزامنة. و نتيجة لعدم تجانس الأطراف المتصلة تضاف خاصية الاستقلال الدلالي كبعد رابع، ولكن هذه الاضافة عقدت عملية المطابقة و خفضت كفاءتها، لذلك تم جمع الاشتراكات والاحداث بشكل عناقيد تبعا للمواضيع لإنجاز عملية المطابقة ضمن تلك العناقيد . حيث تبين أن عملية الجمع ضمن عناقيد أدت لزيادة الانتاجية نتيجة تخفيض زمن و دقة المطابقة عندما تصل الاشتراكات الى حدودها القصوى . وبالتالي إظهار فوائد العناقيد بالإضافة إلى تحسين دقة و كفاءة المطابقة عبر استخدام هذا النهج .

الكلمات المفتاحية: انترنت الأشياء، أنظمة النشر والاشتراك، الارتباط الدلالي ، مطابقة الحدث .

• أستاذ، قسم هندسة الاتصالات والالكترونيات ، كلية الهندسة الميكانيكية والكهربائية، جامعة البعث، حمص سوريا. assalem1@gmail.com

•• طالب دراسات عليا (دكتوراه) ، قسم هندسة الاتصالات والالكترونيات ، كلية الهندسة الميكانيكية والكهربائية، جامعة البعث، حمص سوريا. wael-habeeb@live.com

مجلة جامعة تشرين للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (40) العدد (6) 2018
Tishreen University Journal for Research and Scientific Studies - Engineering Sciences Series Vol. (40) No. (6) 2018

Developing IOT Environment by Using Publish/Subscribe Systems

Dr. Abdulkarim Assalem •
Wael Habeeb ••

(Received 23 / 9 / 2018. Accepted 26 / 12 / 2018)

□ ABSTRACT □

IOT sensors use the publish/subscribe model for communication to benefit from its decoupled nature with respect to space, time, and synchronization. Because of the heterogeneity of communicating parties, semantic decoupling is added as a fourth dimension. The added semantic decoupling complicates the matching process and reduces its efficiency. The proposed algorithm clusters subscriptions and events according to topic and performs the matching process within these clusters, which increases the throughput by reducing the matching time . Moreover, the accuracy of matching is improved when subscriptions must be fully approximated . This work shows the benefit of clustering, as well as the improvement in the matching accuracy and efficiency achieved using this approach.

Key words : Internet of Things (IOT) **,** Publish-Subscribe ، Semantic coupling , Event matching .

• **Professor, Department of Communication and Electronics, Faculty of mechanical and electrical engineering, AL-Baath University, Homs, Syria. assalem1@gmail.com**

•• **PhD student, Department of Communication and Electronics, Faculty of mechanical and electrical engineering, AL-Baath University, Homs, Syria. wael-habeeb@live.com**

مقدمة :

ازداد الاهتمام مؤخراً بموضوع انترنت الاشياء بشكل ملحوظ ليحتل عناوين البحث الرئيسية وذلك لفوائده العديدة وتطبيقاته المهمة كالمدن الذكية و المنازل الذكية و الاستشعار البيئي و ادارة المخلفات. حسب تقرير (غارتنر) الصادر في عام 2014 و 2015 [1,2] فإن ما يقارب 4.9 مليار شيء متصل سيكونون قيد العمل بحلول عام 2015 و 1.1 مليار منهم سيدخلون المدن الذكية. وقد تجاوزت الارقام الواقعية آنذاك تلك التوقعات، كما توقع التقرير تزايد هذا العدد ليصل إلى 25 مليار بحلول عام 2020. هذا التزايد المضطرد في أرقام حساسات الاتصال أوصلنا لكميات غير متوقعة لحركة البيانات و الأحداث التي تحتاج للتواصل بكفاءة قدر الامكان. فباتت عملية ادارة البيانات الضخمة بالطرق التقليدية و تقنيات التخزين كنموذج الهيكلة والتجميع ....غير فعالة لحساسات انترنت الاشياء لان الاتصال الزمني المعتمد على معالجة الاحداث مهم في انترنت الاشياء. يمكن أن يلعب النشر\الاشتراك القائم على المعدات الوسيطة دورا هاما في هذه الادارة [3,4,5]. حيث يعتمد نمط اتصاله على فصل التفاعل بين منتجي المعلومات و مستهلكيها.

يقدم النشر\ الاشتراك المقاد بالمعدات الوسيطة نموذجاً للاتصال الملائم لأنه يؤمن عملية الفصل في أبعاد المزامنة و الزمن و المكان وذلك بأقل استخدام ممكن وهذا ما سمح له بأن يكون أحد النماذج الفعالة للاتصال ضمن انترنت الاشياء ليتجاوز العديد من المشاكل ضمن تلك البيئات. نشرت مجموعة البحث الاوروبية عن انترنت الاشياء تقريرا في آذار 2015 عن أهمية دور انترنت الاشياء و تحدياته البحثية. حيث أكد التقرير على الحاجة لمعدات النشر\ الاشتراك الوسيطة الدلالية [3]. حتى أنه اقترح بأن يتم فك أو تخفيف الارتباط الدلالي في نموذج النشر\الاشتراك للاتصال ليلائم بيئة انترنت الاشياء.

يقصد بفك الارتباط : إلغاء الصلة بين ناشري المعلومات والمشتركين بها عن طريق إزالة التبعيات الصريحة بينهم.

بالنظر إلى الأبعاد المشتركة لفك الارتباط في نموذج النشر / الاشتراك ، يمكن للمشاركين التواصل دون معرفة بعضهم البعض في أوقات مختلفة ، ويجب عدم حظر أي منهم من الاتصال [9] .

إن معظم أنظمة النشر / الاشتراك تفترض اتفاقاً على نحو واسع ، وهو تطابق دقيق للأحداث والاشتراكات بين جميع الأطراف المتعاملة . حيث يرتبط الاتصال بمخطط ومضمون واضح المعالم ومتفق عليه.

ولكن كفكرة أساسية فإن الاتصال بين مزودي ومستهلكي البيانات العديدين والغير متجانسين بالنظر إلى خلفياتهم المتنوعة ، لا يتطلب اتفاقاً على شروط محددة مسبقاً [10 , 6] ، مع ذلك فقد أجريت دراسات قليلة على إضافة الفصل الدلالي (semantic decoupling) [6, 10] داخل نموذج النشر / الاشتراك . ولكن هذا النهج يواجه العديد من التحديات مثل مطابقة الاحداث مع الاشتراكات المناسبة نظراً لعدم وجود اتفاق على مخطط الحدث .

التحدي الآخر هو التسليم الفعال للأحداث نظراً للعدد الكبير للمشتركين الغير متجانسين [11] بالإضافة إلى أن البرمجيات الوسيطة الدلالية (middle ware) . ونقصد بمصطلح (قابلة للتوسع) بأن تلك البرمجيات يجب أن تكون قادرة على استيعاب البيانات الكبيرة الناتجة عن العدد الكبير من أجهزة الاستشعار وأحداث التواصل.

نظراً للحاجة إلى الكفاءة في مطابقة أحداث النشر / الاشتراك مع الأخذ بعين الاعتبار لتعقيد مطابقة البيانات المرتبطة دلالياً أكثر من المحتوى الدقيق ، يتم تلخيص الطريقة كما يلي :

1- تم افتراض بنية عنقودية الشكل تقوم بتجميع الأحداث والاشتراكات وفقاً لموضوعات مميزة ، ثم تم إجراء مطابقة في الأحداث والاشتراكات الواقعة ضمن العناقيد المتقاطعة.

2- تم افتراض خوارزمية تقوم بمطابقة الاحداث والاشتراكات ضمن تلك العناقيد والقيام باختبار هذه الخوارزمية على مجموعة بيانات متنوعة واقعية .

3- تم اجراء تجارب على الخوارزمية المقترحة لتحديد مدى دقتها والانتاجية.

1-1 بنية نظام النشر والاشتراك :

تم تقديم نموذج اتصالات النشر / الاشتراك في التسعينيات من قبل Birman & Joseph [12] ومنذ ذلك الحين أثبت هذا النموذج كفاءته ليكون أسلوباً ناجحاً وواسع الانتشار في الأنظمة الكبيرة الموزعة .

يعرف نظام النشر/الاشتراك على أنه عبارة عن نظام اتصال، يتم فيه تبادل المعلومات على شكل أحداث. حيث يقوم الناشرون بتسليم المعلومات (الأحداث) إلى خدمة تسليم الحدث التي تقوم بتسليمها إلى المُشتَركين المهتمين بتلك الأحداث .

الشكل (1) بنية بسيطة لنظام نشر و اشتراك

2-1 تطبيقات أنظمة النشر والاشتراك :

عند الاطلاع على التطبيقات التي تستخدم أنظمة النشر والاشتراك ندرك أهمية تلك الأنظمة كونها تعدت مجالات الترفيه والتواصل الاجتماعي لتدخل المجال الاكاديمي والاقتصادي والصحي ..... ومن تلك التطبيقات :

2-1-1 الأنظمة المالية Financial Systems :

تعتبر سوق الأسهم المالية Stock Market من أبرز الأمثلة عن الأنظمة المالية التي تهتم بتمكين المتعاملين من الحصول على آخر المعلومات عن أسعار الأسهم التي يتعاملون بها. والشكل (2) يقدم نموذجاً مبسطاً لسوق الأسهم المالية.

الشكل (2) بنية بسيطة لنظام نشر و اشتراك

2-1-2 الألعاب التفاعلية ذات العدد الهائل من اللاعبين Massively Multiplayer Online Games :

في هذا التطبيق ، عندما يقوم أحد اللاعبين بأداء فعل معين يؤثر على مجرى اللعبة، يجب أن يحصل بقية اللاعبون على تحديثات تتعلق بالتغيرات التي نتجت عن ذلك الفعل. الشكل (3)

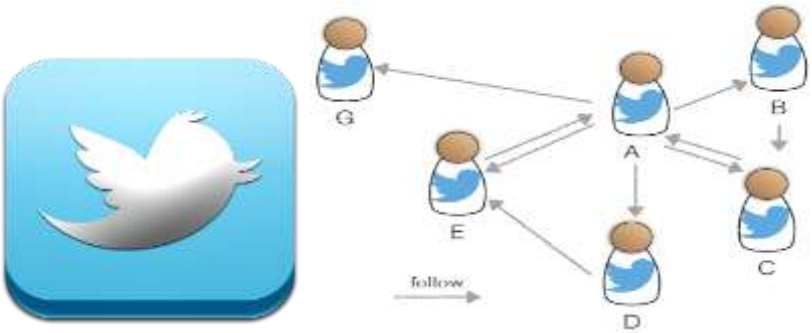
الشكل (3) الالعاب التفاعلية

2-1-3 تويتر Twitter :

خدمة شائعة من خدمات التواصل الاجتماعي عبر الأنترنت.

- كل رسالة تدعى تغريدة Tweet .

- بينت الدراسات أن هناك 500 مليون مستخدم مسجل في هذه الخدمة في عام 2013، وكان أكثر من 200 مليون من هؤلاء المستخدمين المسجلين نشط يومياً . الشكل (4).



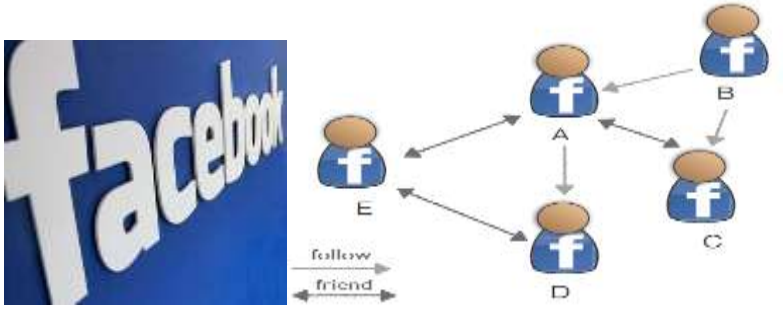
الشكل (4) : مثال عن العلاقات بين المستخدمين في تويتر

2-1-4 فيس بوك Facebook :

يعتبر أهم مواقع التواصل الاجتماعي على الأنترنت أهمية وانتشاراً.

- إن العلاقات في الفيس بوك ثنائية الاتجاه بشكل عام. يعني بأنه على المستخدمين أن يقبلوا صداقة كل منهم الآخر ليكونوا أصدقاء.

- بالإضافة لعلاقة الصداقة فإن موقع الفيس بوك يسمح بالعلاقة أحادية الاتجاه على غرار موقع تويتر. الشكل (5).



الشكل (5) : مثال عن العلاقات بين المستخدمين في فيسبوك

أهمية البحث وأهدافه :

تحتاج الرسائل بين الكائنات المختلفة في بيئة مفتوحة على نطاق واسع مثل انترنت الأشياء إلى تجاوز حدود النظام . والحدين الرئيسيين هما : الحد النحوي (syntactic) ، والحد الدلالي (semantic).

1- الحد النحوي (syntactic): يهتم بقضايا مثل تنسيق البيانات والعنونة .

2- الحد الدلالي (semantic): يظهر عندما تستخدم أنظمة مختلفة مصطلحات مختلفة لمفاهيم متشابهة أو ذات صلة. وهذه الحدود الدلالية لا مفر منها في الأنظمة الكبيرة الغير متجانسة .

يتجاوز الحد النحوي للنظام مخطط النشر / الاشتراك المعتمد على المحتوى والموضوع، وفيه يتفق المشتركون والناشرون على شروط محددة وعدد كبير من القواعد الموضوعة مسبقاً من أجل المفاهيم المختلفة التي سيتم استخدامها في الاتصال. مما يعني بأن ارتباط الطرفيات باتفاقات مسبقة ضروري لتحقيق نقل أحداث فعال ضمن نظام انترنت الاشياء . ويعرف الارتباط الدلالي (semantic coupling) حسب [8] بأنه : مقدار الاتفاقيات بين المشتركين في بيئة معالجة الحدث في الارتباطات بين الرموز المستخدمة في رسائل الاحداث والمعاني التي تشير إليها ، ولحد الآن لا يعتبر الارتباط الدلالي الوثيق ضمن بيئة انترنت الاشياء أمراً واقعياً لأن هذه البيئة تعتبر قابلة للتوسع بشكل كبير ومفتوحة وغير متجانسة ، حيث تعتبر مطابقة الأحداث للاشتراكات في معالجة الاحداث الدلالية التقريبية أمر تقريبي وليس منطقي مثل طرق الأحداث المعتمدة على المفهوم والتي تكون محددة بواسطة مجال معين مسبقاً ، وبالتالي فإن عدم اليقين أو عدم الموثوقية أو البيانات الخاطئة أو عدم الاكتمال، جميعها أمور لا مفر منها في انترنت الاشياء ، نظراً لأن الاتصال في انترنت الاشياء يكون غالباً بين الآلات أو البشر والآلات والتي قد تختلف في مستوى موثوقيتها . [22]

في هذا البحث استخدمنا الدلالة التقريبية بدلاً من القيمة المنطقية الدقيقة في التواصل بين الطرفيات المختلفة لتكون أكثر مرونة وملائمة لطبيعة هذا التواصل عبر صياغة أسئلة البحث التالية:

1- كيف نستطيع تحقيق المطابقة الدلالية بدلاً من تحديد عملية المطابقة بخيارات محددة عبر استخدام أقل التفاعلات أو المدخلات من المستخدمين أو الحساسات؟

2- كيف يمكن لهذا النظام أن يكون فعالاً بالنظر إلى التعقيدات المضافة للتقليل من الارتباط الدلالي؟

وبالتالي سيكون التركيز منصباً في البحث على تعزيز مخطط النشر / الاشتراك مع المحافظة على الاستخدام السهل والبسيط للاتصالات.

3- خصائص ومميزات أنظمة النشر/الاشتراك :

تتميز أنظمة النشر/الاشتراك بالصفات التالية :

1- عدم الارتباط المكاني : الناشر والمشترك مستقلين عن بعضهما مكانيا فيمكن أن يكون المشترك في دولة والناشر في دولة أخرى .

2- عدم الارتباط الزماني : الناشر والمشترك ليسا بحاجة أن يكونا فعالين في نفس الوقت لاستقبال وإرسال المعلومات .

3- عدم الارتباط من ناحية التزامن : تعني أن المشترك عمله لا يتوقف على الناشر والعكس بالعكس .

- إن طبيعة الاتصال غير المرتبط في تلك الشبكات هي نفسها الطريقة الأساسية للاتصالات في الأنترنت ، وبالتالي ضمن بيئة انترنت الأشياء (IOT) [9].

4- التصنيفات الأساسية لشبكات النشر / الاشتراك :

4-1 من حيث طريقة تسجيل الاشتراك :

يكون التسجيل فيها اعتمادا على الموضوع أو على المحتوى

4.1.1. الطريقة الاولى حسب الموضوع Topic-Based [13,14,15,16] :

يسجل المشتركون فيها اعتمادا على كلمة مفتاحية تمثل اسم الموضوع أي يعلن المشترك رغبته في موضوع محدد يختاره كأخبار الرياضة مثلا.

4.1.2. الطريقة الثانية حسب المحتوى Content-Based [17,18] :

يسجل المشتركون فيها اعتمادا على محتوى الموضوع أي يتم التركيز على كلمة ضمن الموضوع مثل تسجيل الاشتراك بالأخبار التي تتحدث عن عالم مشهور .

- إن التصنيفات التي تقدم ذكرها تعتمد على مطابقة قيمة دقيق أو ضمن مجال تقريبي وكلا الطرفين لا يأخذان بعين الاعتبار احتمالية استخدام مصطلحات مختلفة لنفس الكلمة أو المفهوم . لذا في البيئات المتباينة حيث أطراف الاتصال لا تشترك بنفس الخلفية، سيتم فقد العديد من المطابقات . وهكذا ظهر صنف جديد من خوارزميات النشر / الاشتراك والذي يأخذ بعين الاعتبار مصطلحات المطابقة المرتبطة دلالياً .

4.2. من ناحية التطابق الدلالي :

يطابق نموذج النشر / الاشتراك الدلالي الاشتراكات مع الاحداث ذات الصلة وفقاً للمعنى المنطقي. حتى الآن ، يستند نموذج النشر / الاشتراك التقليدي على مفهوم عدم الارتباط من ناحية الزمن والمكان والتزامن. وبالمقابل فإن الارتباط الدلالي يكون وثيقاً فيما يتعلق بالاشتراكات والصفات والعوال الواضحة المستخدمة لوصف الدلالات الكامنة وراء الاحداث المتبادلة ، بينما يفترض الناشرون والمشتركون التقليديون اتفاقاً على أنواع الحدث وخصائصه وقيمه.

وفيما يلي التصنيفات البديلة لخوارزميات المطابقة الدلالية للنشر / الاشتراك :

4.2.1. الطرق المعتمدة على المفهوم (concept-based methods):

تعتبر المطابقة ضمن نهج الاعتماد على المفهوم عبارة عن مطابقة دلالية منطقية تعتمد على اتفاق مسبق لمستوى فهم مشترك معين أو استخدام مجال علمي محدد [19] [S-Topps] كان أول من قدّم مفهوم المطابقة الدلالية من خلال استخدام التسلسل الهرمي لكل مفهوم مثل التعميم والتخصيص أو وظائف رسم المخطط التي تسمح بعلاقات كيفية بين المخطط وقيم السمات لخوارزميات النشر / الاشتراك .

- في [20] استخدم الباحثون إطار عمل وصفي مصدري (resource description framework) (RDF) لتقديم الأحداث بدلاً من مقارنة الخاصية والقيمة.

- في أنظمة النشر / الاشتراك المعتمدة على المفهوم ، تتم عملية المطابقة باستخدام العلوم الموجودة (Ontologies) أو مجموعة مفاهيم مشتركة متّفق عليها والتي تمنع فك ارتباط المصطلحات . وعلاوة على ذلك فإن إضافة مصطلح جديد يمكن أن يصبح مرهقاً لما يسببه من بطء .

4.2.2. مطابقة الحدث الدلالي التقريبي :

تعتمد مطابقة الحدث الدلالي التقريبي على نموذج إحصائي مبني اعتماداً على دلالات ألفاظ مصنفة وموزعة . حيث تنظم النتائج اعتماداً على القيم التقريبية .

يعتبر [21] [A-Topps] أول بحث يقدم مفهوم المطابقة التقريبية (approximate matching) عبر استخدام معاملات تقارب بدلاً من المعاملات الصريحة المحددة التي تستخدمها خوارزميات النشر / الاشتراك .
كما أن المرجع قام بتقديم مفهوم جديد يدعى درجة الارتباط (match degree) بدلاً من القيمة المنطقية المحددة ب (نعم) أو (لا) . هذا المفهوم الجديد يستخدم نظرية الاحتمال لتحديد درجة التقارب.
لاحقاً في [7] تم القيام بعمل رئيسي حيث قدم الباحثون الأحداث كإطار عمل وصفي مصدري (RDF) ، والاشتراك كدمج للبيانات أو الحالات التي تمت ترجمتها إلى نمط لتكون بمثابة استعلام عن الأحداث ، حيث اعتمدت المطابقة على الأحداث المرتبطة دلالياً والتي تم تصنيفها اعتماداً على التشابه. وبالتالي فإن الناتج كان رقماً حقيقياً.

4.2.3. معالجة الأحداث ذات الصلة (Thematic Event Processing) :

في هذا التصنيف يتم إضافة علامات رئيسية لزيادة سرعة عملية المطابقة . في [8,10] قام الباحثون بإضافة علامات رئيسية تمثل الاشتراك والمجال ، وبالتالي تم تصفية المطابقة وفقاً للعلامات التي تطابق الاشتراكات المرغوبة ، وبالتالي تزداد الانتاجية .
مما سبق نجد بأن البرمجيات التي تعتمد على الحدث في شبكات النشر / الاشتراك مناسبة جداً في الاتصالات بين الحساسات في شبكات انترنت الأشياء (IOT) بسبب طبيعتها وخصائصها في عدم الارتباط الزماني والمكاني وفي التزامن .
بالنظر إلى ما وصلت إليه تطبيقات انترنت الأشياء من تقدم كبير مثل المدن الذكية والمنازل الذكية والمستشعرات عن بعد ، والتي تعمل على نطاق واسع للغاية باستخدام بيانات كبيرة الحجم. هذا التواصل واسع النطاق بين الطرفيات غير المتجانسة يزيد من الحاجة إلى فك الارتباط الدلالي ، حيث تم إجراء بعض البحوث لمعالجة هذه الحاجة [6,8,19,21] ، ولكن عند إضافة فك الارتباط الدلالي فإن نظام النشر / الاشتراك يصبح أكثر تعقيداً ، مما يؤثر سلباً على كفاءته .

طرائق البحث ومواده:

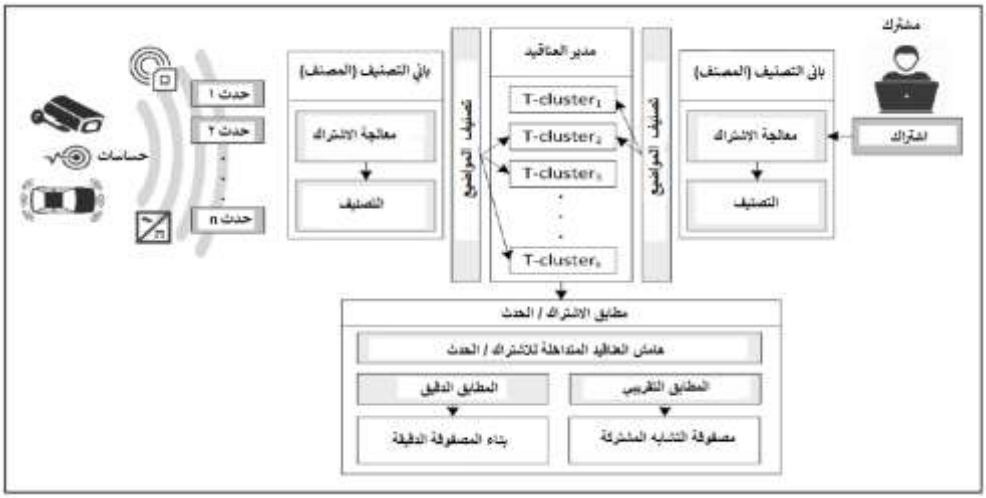
تم تطبيق خوارزمية (CURRY & HASAN) [6] والتي تعتمد على نظام المطابقة المنطقية الدقيقة للتقارب الدلالي في المقارنة بين المشتركين والناشرين، وقارنا النتائج مع النظام المقترح في هذا البحث من خلال استخدام برنامج PeerSim الخاص بنمذجة الشبكات والذي يعمل ضمن البيئة البرمجية eclipse وفيما يلي النتائج التجريبية المحسوبة على جهاز (DELL) بالمواصفات التالية:
Intel core i5 , 3.4-GHZ cpu with 4 GB of RAM with Microsoft windows 7 , 64- bit operating system.

6- بنية النظام المفترض :

يمثل الشكل (6) النظام المفترض ، حيث يأخذ الأحداث والاشتراكات كمدخلات إلى باني التصنيفات (المصنف) . يقوم هذا المصنف بمعالجة الاشتراكات والأحداث ، ثم يصنفها. تتوضع الأحداث والاشتراكات للتجميع تحت عناقيد المواضيع التي تماثلها.

الأحداث التي تقع ضمن كل عنقود اشتراك يتم مطابقتها إما بشكل دقيق أو بشكل تقريبي مع توقعات الاشتراك كما هو مطلوب، ويتم بناء مصفوفة وفقاً لذلك للمطابقات الدقيقة والمطابقات التقريبية .

يتم معالجة المصفوفات الناتجة كما سيتم شرحه في القسم التالي.



الشكل (6) بنية نظام مطابقة الحدث

7- الخوارزمية المفترضة :

7.1. تطبيق التحليل الدلالي الصريح (Explicit Semantic Analysis) (ESA) :

يتم استخدام مطابقة تقريبية في الخوارزمية المفترضة ، حيث نعتمد على التحليل الدلالي الصريح (ESA) [23,24] والذي أثبت أنه يسفر عن نتائج جيدة لقياس الارتباط الدلالي .

علاوة على ذلك، يستخدم (ESA) لحساب التشابه في مدونة (Wikipedia) ، والتي تناسب بشكل كبير تطبيقات انترنت الأشياء ، لأنها تتوسع بشكل دائم من قبل المستخدمين وهي غير محدودة بشكل مسبق .

تم تحميل ملف تفريغ Wikipedia لعام 2016 [25] ، و لفهرسته تم بناء جدول تشير فيه الأعمدة إلى المفاهيم التي تتعلق بمقالة أو مستند معين (d) ، وتشير الصفوف إلى الكلمات أو المصطلحات (t) التي توجد في المقالة .

- يمثل كل إدخال في جدول قيمة (term frequency–inverse document frequency)(TFIDF) للمصطلح ti في المستند di حيث تمثل (TFIDF) قيمة تردد المصطلح ومقلوب تردد المستند.

لحساب (TFIDF) نقوم بحساب المكونات التالية :

- تردد المصطلح (tf) :

$$\mathrm{tf}(t_i , d_j) = \begin{cases} \mathrm{count}(t_i , d_j) & \text{if count}(t_i , d_j) > 0 \\ 0 & \text{otherwise} \end{cases}$$

المعادلة (1) تردد المصطلح

- تردد المستند (df) :

هو عدد المستندات التي تحوي المصطلح ti ضمن مجموعة المستندات الكلية المختارة .

$$df_i = |\{dk : t_i \in d_k\}|$$

المعادلة (2) تردد المستند

– مقلوب تردد المستند (IDF) :

$$(idf_i = \log \frac{n}{df_i})$$

المعادلة (3) مقلوب تردد المستند

حيث (n) تمثل عدد المستندات ضمن المجموعة .

$$T[i,j] = \frac{T[i,j]}{\sqrt{\sum_{i=1}^{k} T[i,j]^2}}$$

المعادلة (4) عدد المستندات

حيث k تمثل عدد المصطلحات .

فيما يلي الجدول(1) الذي يشير إلى قيم (TFIDF) المعبرة عن صلة المفهوم بالمصطلح ، حيث تعبر الصفوف عن المصطلحات والأعمدة عن المستندات.

الجدول (1) جدول قيمة (TFIDF)

	مستند 1	J	مستند n
مصطلح 1	T [ti , dj] = TFIDF		T [ti , dj] = TFIDF
i			
مصطلح n	T [ti , dj] = TFIDF		T [ti , dj] = TFIDF

7.2. تخصيص الموضوع عبر التصنيف :

لتقليل حجم مساحة البحث عند إيجاد مطابقة ذات صلة ، فإن الأحداث والاشتراكات توضع في عناقيد منفصلة . ولتحقيق ذلك اعتمدنا طريقة مستوحاة من حقيقة أن أنظمة المطابقة في شبكات النشر والاشتراك المعتمدة على الموضوع أكثر كفاءة.

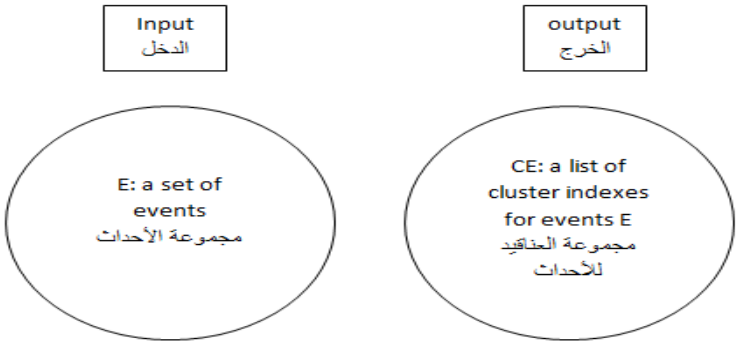
نقوم بتعيين موضوع لكل حدث او اشتراك . يتم مطابقة الأحداث ضمن العنقود المرتب بنفس الموضوع. والسبب وراء ذلك بأن بيئة انترنت الأشياء متنوعة للغاية بطبيعتها، ولذا فإن تخصيص موضوع للاشتراكات والأحداث سيصفّي العديد من الأحداث غير المرغوبة لتجنب الوقت المحسوب في عملية المطابقة.

لتصنيف الأحداث والاشتراكات اعتمدنا على قاعدة بيانات مصنفة إلى (53) مجموعة مختلفة.

7.3. الحدث والاشتراك ضمن العنقود (clustering) :

كل حدث يحوي معرف مميز (ID) ، وهذا المعرف يتوضع ضمن عناقيد القيمة والصفة الخاصة بها ، وبالمثل فإن كل اشتراك يحوي معرف مميز (ID) ، وهذا المعرف يتوضع ضمن عناقيد القيمة والصفة الخاصة بها.

يوضح الشكل التالي الخوارزمية (1) والتي تبين بناء قائمة مجموعة الأحداث . حيث تأخذ الخوارزمية مجموعة أحداث كدخل لها وتقوم بإخراج مجموعة من العناقيد التي يقع كل من هذه الأحداث ضمن إحداها .

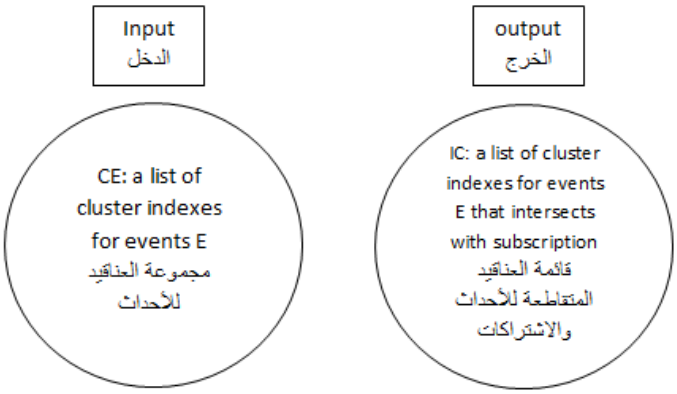


الشكل (7) الخوارزمية (1)

7.4. المطابقة :

يتم مطابقة الاشتراكات مع الاحداث التي تقع ضمن نفس العناقيد . هذه المطابقة السابقة (ضمن نفس العناقيد) هي فقط التي تخفض زمن المطابقة بشكل ملحوظ ، وبالتالي زيادة الفعالية.

وفيما يلي شكل يوضح الخوارزمية (2) التي تبين عملية إيجاد العناقيد المتقاطعة للأحداث والاشتراكات.



الشكل (8) الخوارزمية (2)

الخوارزمية تأخذ كدخل لها قائمة العناقيد للأحداث و المعرف (ID) للاشتراك المعطى.

7.5. بناء المصفوفة :

يتم مطابقة الاشتراكات مع الاحداث في العناقيد التابعة لها.

يبنى أربع بارامترات لكل زوج (اشتراك / نشر) :

1) مصفوفة الصفة الدقيقة = { 0 في حال كان (الاشتراك / الصفة) يتطلب توافق تام وكان الاشتراك / الصفة ≠ الحدث / الصفة

= { 1 عدا ذلك

2) مصفوفة القيمة الدقيقة = { 0 في حال كان (الاشتراك / القيمة) يتطلب توافق تام وكان الاشتراك / القيمة ≠ الحدث / القيمة

= { 1 عدا ذلك

3) مصفوفة الصفة التقريبية = { ESA في حال التوافق التقريبي للصفة بين الاشتراك والحدث

= { 1 عدا ذلك

4) مصفوفة القيمة التقريبية = { ESA في حال التوافق التقريبي للقيمة بين الاشتراك والحدث

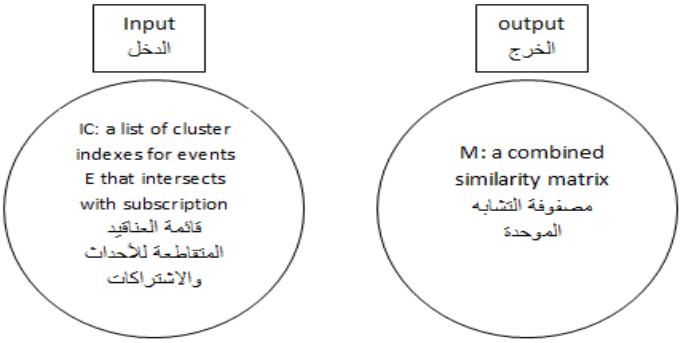
= { 1 عدا ذلك

– قيم المصفوفة التقريبية يمكن أن تكون أي قيمة ضمن المجال [1–0] ، وتلك الأرقام في البارامترات التقريبية تعتمد على نتيجة ESA التي حصلنا عليها بمقارنة صفة أو قيمة الاشتراك مع الحدث.

– جميع المصفوفات الدقيقة والتقريبية هي I*J حيث:

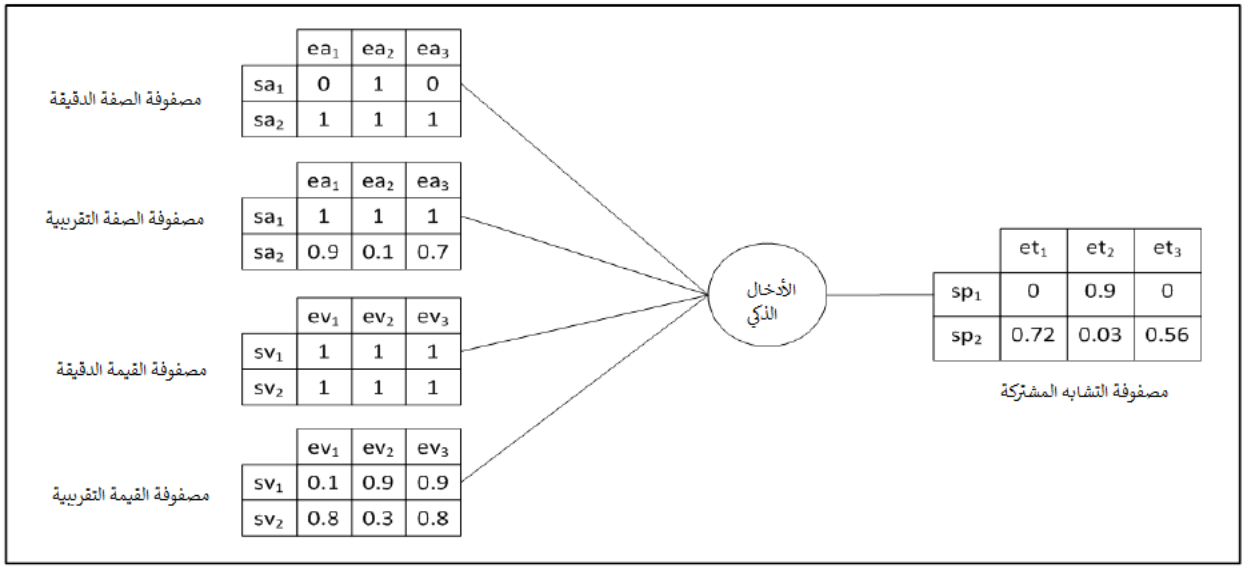
I : عدد المسندات في الاشتراك J : عدد الصفوف التي سيتم مطابقتها في الحدث.

يتم تجميع المصفوفات الأربعة في مصفوفة واحدة باستخدام الإدخال الذكي ، وقيمة واحدة مقنعة تقدم درجة التطابق بين مسند الاشتراك وصف الحدث . الشكل (10) يعطي تصور لكيفية بناء المصفوفة الموحدة كما هو موضح في الخوارزمية (3) .



الشكل (9) الخوارزمية (3)

وفيما يلي الشكل (10) :



	ea_1	ea_2	ea_3
sa_1	0	1	0
sa_2	1	1	1

	ea_1	ea_2	ea_3
sa_1	1	1	1
sa_2	0.9	0.1	0.7

	ev_1	ev_2	ev_3
sv_1	1	1	1
sv_2	1	1	1

	ev_1	ev_2	ev_3
sv_1	0.1	0.9	0.9
sv_2	0.8	0.3	0.8

	et_1	et_2	et_3
sp_1	0	0.9	0
sp_2	0.72	0.03	0.56

الشكل (10) الإدخال الذكي للقيم الدقيقة والتقريبية

8- قاعدة البيانات :

تم اختيار قاعدة بيانات متوفرة للعموم [26] تتضمن أحداث متوائمة مشابهة لتلك التي تم تقديمها في مشروع المدينة الذكية (سانتاندر الذكية) [27] وقاعدة بيانات مشروع الطاقة الذكية المرتبطة (Linked Energy Intelligence) (LEI) [28] .

(سانتاندر الذكية) هو مشروع قياس مدينة ذكية مكون من (3000) حساس موضوعة في مدن اوروبية محددة في أماكن مختلفة مثل محطات توقف النقل العام ، الكراجات ، الباصات وسيارات الأجرة، لمراقبة جوانب مختلفة مثل الازدحام ، امكانية إيقاف السيارات ، ومعايير بيئية معينة مثل التلوث . بينما تعتبر قاعدة بيانات (LEI) كمشروع بناء ذكي ركّز على توفير الطاقة وبعض المعايير المرتبطة بذلك.

النتائج والمناقشة :

تم تقييم أداء نظام مطابقة الحدث المفترض لإثبات فعاليته وكفاءته.

* لاختبار كفاءة الخدمات التي يقدمها نظامنا ، استخدمنا البارامتر التالي:

- الإنتاجية (Through-put) عدد الأحداث المتطابقة في زمن محدد.

* لاختبار الفعالية لنظامنا المفترض تم استخدام القياسات التالية:

- الدقة (precision) تقيس نسبة الأحداث المرتبطة المكتشفة من قبل المطابق مع الأخذ بعين الاعتبار لجميع الأحداث المكتشفة.

- الاسترجاع (Re call) تقيس نسبة الأحداث المرتبطة المكتشفة من قبل المطابق مع الأخذ بعين الاعتبار لجميع الأحداث المرتبطة من قاعدة البيانات .

- (F-score) يضم الدقة واعادة الاستدعاء سوياً بشكل متساوي، يحسب بالعلاقة :

$$F\text{-}Score = (2 * Precision * Recall)/ (Precision + Recall)$$

المعادلة (5) حساب قيمة F-score

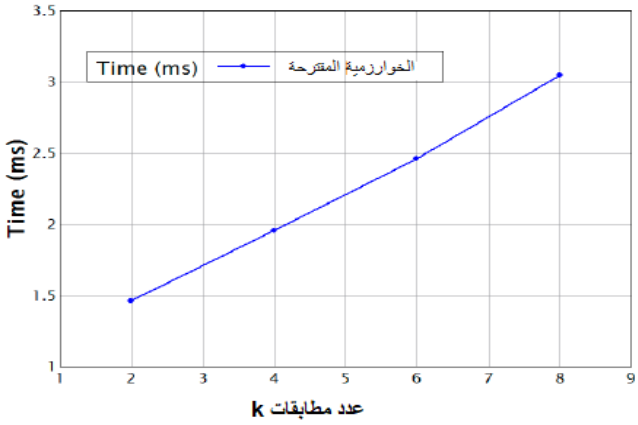
1.. الكفاءة : الإنتاجية (Efficiency: Throughput):

يبين الشكل (11) الانتاجية لنظام مطابقة الحدث والذي يعطي عدداً مختلفاً من مطابقات (top-k) بالنسبة لعدد ثابت من مسندات الاشتراك (n) والذي يساوي عدد صفوف الأحداث (m) بحيث:

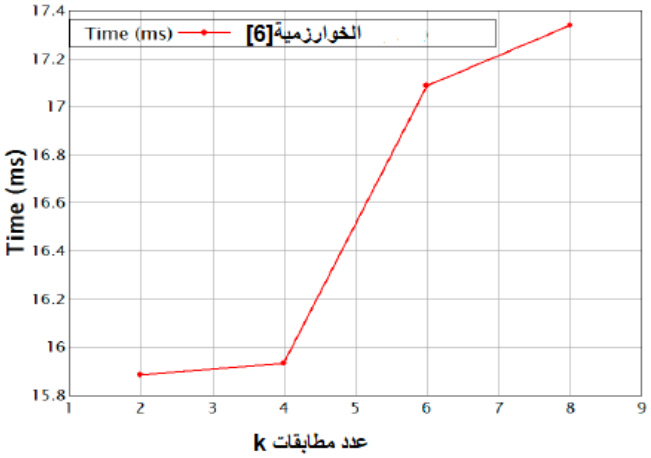
$$m = n = 10$$

نلاحظ أن الخوارزمية المفترضة خطية تقريباً مع (k) واسرع من خوارزمية (CURRY & HASAN) [6] من أجل جميع قيم (k) .

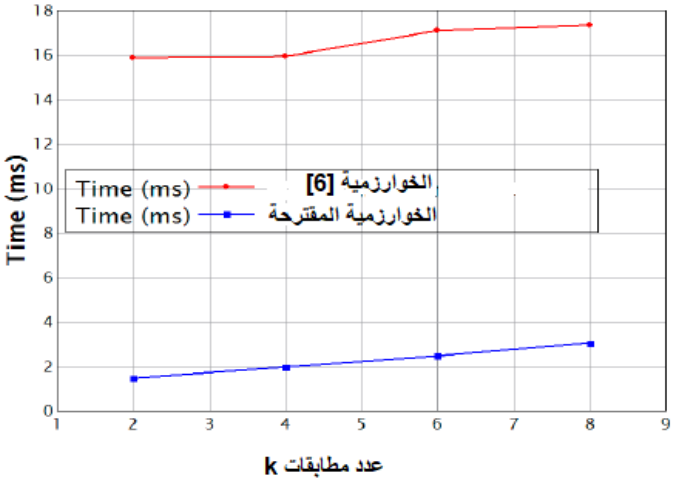
- الشكل (12) يبين انتاجية المطابقة من أجل عدد متغير لمسندات الاشتراك . كلا الشكلين يوضحان أن خوارزمية (CURRY &HASAN) تحتاج وقتاً أقل عندما يكون عدد المسندات صغيراً. ولكن عندما يزدادا عدد المسندات ، فإن الانتاجية في خوارزميتنا تزداد ، والوقت الإضافي الذي تحتاجه يزداد ببطء أكبر ، وذلك بسبب آلية التجميع ضمن عناقيد (clustering). أي أنه بزيادة عدد المسندات تصبح عملية المطابقة أسرع .



(a)

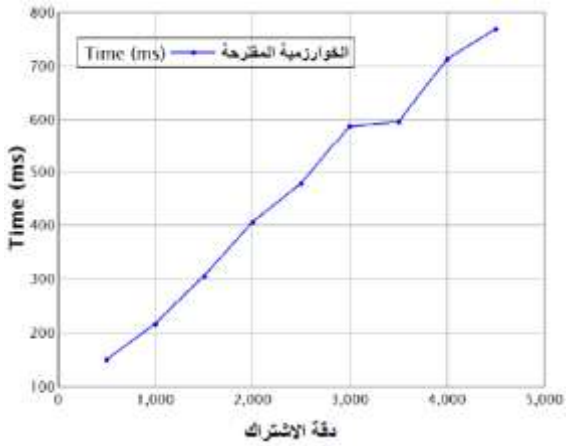


(b) تم أخذ ترقيم الزمن هكذا لسهولة المقارنة بين المنحنيين في الشكل c التالي

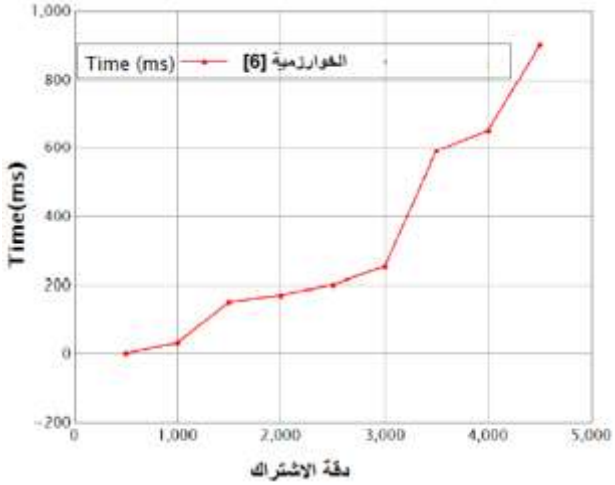


(c)

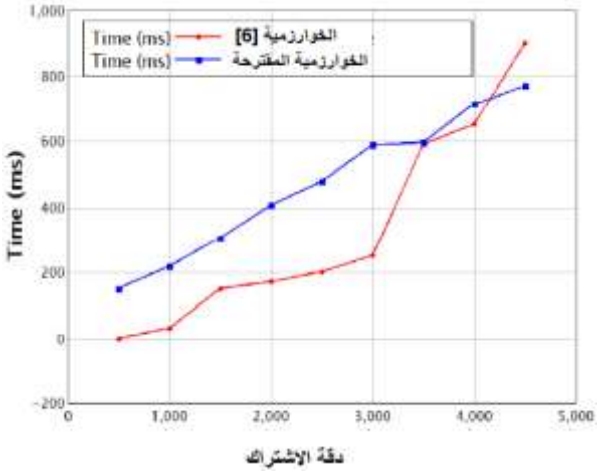
الشكل (11) مقارنة الزمن مع K من أجل الخوارزمية المقترحة (a) ، من أجل الخوارزمية [6] (b) ومن أجل كلا الخوارزميتين معاً (c).



(a)



(b)



(c)

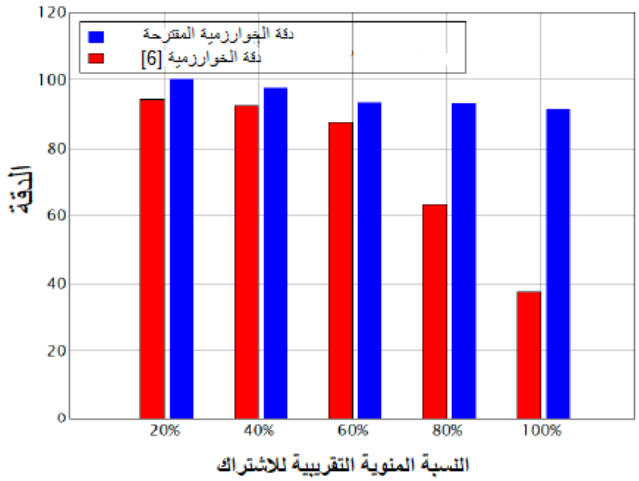
الشكل (12) مقارنة الزمن مع دقة الاشتراك من أجل الخوارزمية المقترحة (a) ، من أجل الخوارزمية [6] (b) ومن أجل كلا الخوارزميتين معاً (c).

2.. الفعالية (effectiveness):

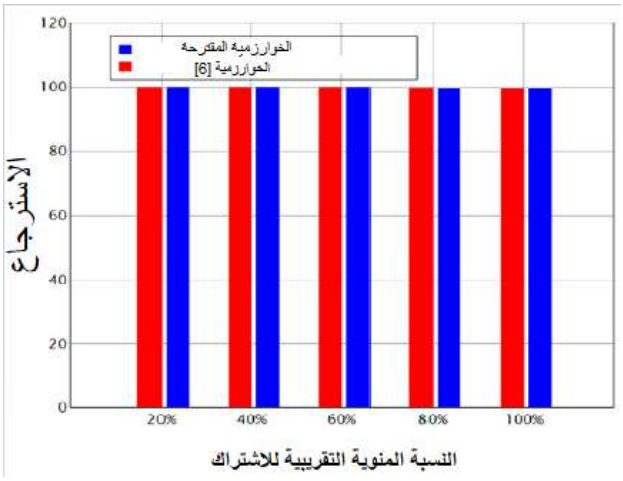
يظهر الشكل (13) النسب المئوية للدقة (precision) ، الاسترجاع (Re call) ، و الـ (F-score) للاشتراكات المتقاربة لخوارزميتنا بالمقارنة مع (CURRY & HASAN) [6] .

نلاحظ في الأشكال التالية ما يلي :

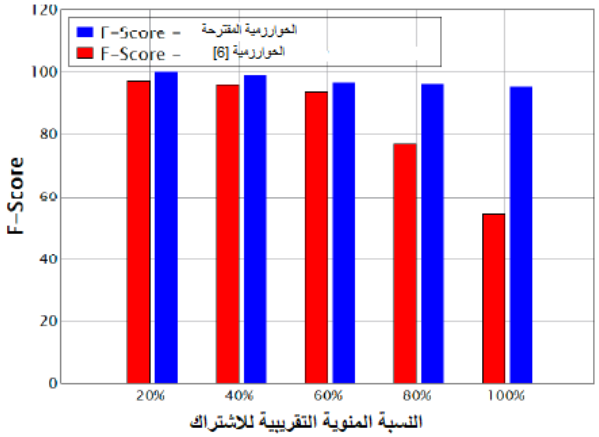
- من أجل نسبة تقريب 80% وما فوق تنخفض دقة خوارزمية (CURRY & HASAN) [6] بشكل حاد ، في حين حافظت خوارزميتنا على الدقة المثالية وتنخفض من 93% إلى 91% .
- قيمة الاسترجاع (Re call) متساوية تقريباً في كلا الخوارزميتين .
- تحقق خوارزميتنا فعالية أكبر مع مقايضة كبيرة بين الدقة والاسترجاع موضح من قبل (F-score) . بينما خوارزمية (CURRY & HASAN) [6] غير قادرة على تحقيق ذلك بسبب العدد الكبير من النتائج الخاطئة للمطابقة الإيجابية .



(a)



(b)



(c)

الشكل (13) مقارنة الفعالية المقاسة في الخوارزمية المقترحة مع الخوارزمية [6]، (a) مقارنة الدقة ، (b) مقارنة الاسترجاع، (c) مقارنة (F-score).

الاستنتاجات والتوصيات :

– إضافة المطابقة الدلالية لنموذج النشر/ الاشتراك يزيد المرونة في الاتصالات ولكن يعقد المطابقة ويخفض الكفاءة .

– في هذا البحث ، تم اقتراح طريقة لتطبيق نموذج النشر / الاشتراك الدلالي بكفاءة من خلال تجميع أحداث واشتراكات متنوعة في عناقيد منفصلة يتم تصنيعها حسب الموضوع، وهذه العناقيد توفر الوقت الذي تحتاجه عملية المطابقة البطيئة بين الأحداث والاشتراكات في خوارزمية .

– هذا يزيد فعالية المطابقة بشكل كبر مع أخذ الدقة (precision) و (F-score) بعين الاعتبار.

المراجع:

[1] RIVERA,J;VANDER,M ; GARTNER,R , http://www.gartner.com/newsroom/id/2905717 , (accessed on 31 APRIL 2018).

[2] RIVERA,J;VANDER,M ; GARTNER,R , http://www.gartner.com/newsroom/id/3008917 , (accessed on 31 APRIL 2018).

[3] SERRANO, M.; BARNAGHI, P.; CARREZ, F.; COUSIN, P. " *Internet of Things: IoT Semantic Interoperability: Research Challenges* " http://www.internet-of-things-research.eu/pdf/IERC_Position_Paper_IoT_Semantic_Interoperability_Final.pdf (accessed on 2 September 2017).

[4] FOX,G.; KAMBURUGAMUVE,S.; HARTMAN,R.D. *"Architecture and measured characteristics of a cloud based internet of things*". In Proceedings of the 2012 International Conference on Collaboration Technologies and Systems (CTS), Denver, CO, USA, 21–25 May 2012; pp. 6–12.

[5] STOPLE,M. *"The Internet of Things: Opportunities and Challenges for Distributed Data Analysis"*. Newsl.2016, 18, pp.15–34.

[6] HASAN, S.; CURRY, E." *Approximate Semantic Matching of Events for the Internet of Things. ACM Trans."* Internet Technol. 2014, pp.14.

[7] HASAN, S.; O'RIAIN, S.; CURRY, E. *"Approximate Semantic Matching of Heterogeneous Events."* In Proceedings of the 6th ACM International Conference on Distributed Event-Based Systems(DEBS2012),Berlin, Germany, 16–20 July 2012; pp. 252–263.

[8] HASAN, S. *"Loose Coupling in Heterogeneous Event-Based Systems via Approximate Semantic Matching and Dynamic Enrichment."* Ph.D. Thesis, National University of Ireland Galway, Galway, Ireland, 2016.

[9] EUGSTER, P.T.; FELBER, P.A.; GUERRAOUI, R.; KERMARREC,*" A.-M. The many faces of publish/subscribe"*. ACM Comput. Surv. 2003, 35, 114–131.

[10] HASAN, S.; CURRY,E. *"Thematic Event Processing. In Middleware"*, ACM: New York, , USA, 2014; pp. 109–120.

[11] WANG, T.; TAO, X.; ZHOU, Y.; LI, P.; ZHAO, C.*" A distributed semantic filtering model based on approximate automata for heterogeneous multi-sensor networks."* Int. J. Sens. Netw. 2016, 20, 46–53.

[12] BIRMAN, K.; JOSEPH, T.*" Exploiting virtual synchrony in distributed systems."* In ACM SIGOPS Operating Systems Review; ACM: New York, NY, USA, 1987; Volume 21, pp. 123–138.

[13] CASTRO, M.; DRUSCHEL, P.; KERMARREC, A.M.; ROWSTRON, A.I.T.*" Scribe: A large-scale and decentralized application-level multicast infrastructure"*. IEEE J. Sel. Areas Communication. 2002, 20, 1489–1499.

[14] BIRMAN, K.P. *"The Process Group Approach to Reliable Distributed Computing. Communication"*. ACM 1993, 36, 37–53.

[15] SETTY, V.; VAN STEEN, M.; VITENBERG, R.; VOULGARIS, S.*" Poldercast: Fast, robust, and scalable architecture for P2P topic-based pub/sub"*. In Proceedings of the 13th International Middleware Conference, Montreal, QC, Canada, 3–7 December 2012; Volume LNCS 7662, pp. 271–291.

[16] CHOCKLER, G.; VITENBERG, R.*" SpiderCast: A Scalable Interest-Aware Overlay for Topic-Based Pub/Sub Communication Categories and Subject Descriptors*". In Proceedings of the 2007 Inaugural International Conference on Distributed Event-Based Systems (DEBS '07), Toronto, ON, Canada, 20–22 June 2007; ACM: New York, NY, USA, 2007; pp. 14–25

[17] CARZANIGA, A.; PAPALINI, M.; Wolf, A.L.*" Content-Based Publish/Subscribe Networking and Information-Centric Networking*". In Proceedings of the ACM SIGCOMM Workshop on Information-Centric Networking, Toronto, ON, Canada, 15–19 August 2011; pp. 56–61.

[18] ANADIOTIS, A.C.G.; PATRIKAKIS, C.Z.; VENIERIS, I.S.*" On the Performance Improvement of Gossip Protocols for Content-Based Publish-Subscribe Through Caching*". Comput. Netw. 2013, 57, 3759–3772.

[19] PETROVIC, M.; BURCEA, I.; JACOBSEN, H.-A.*" S-TOPSS: Semantic Toronto Publish/Subscribe System*". In Proceedings of the 29th International Conference on Very Large Data Bases (VLDB '03), Berlin, Germany, 9–12 September 2003; Franklin County, OH, USA, 2003; pp. 1101–1104.

[20] WANG, J.; JIN, B.; LI, J.*" An ontology-based publish/subscribe system*". In Proceedings of the 5th ACM/IFIP/USENIX International Conference on Middleware; Springer Inc.: New York, NY, USA, 2004; pp. 232–253.

[21] LIU, H.; JACOBSEN, H.-A.*" A-TOPSS: A publish/subscribe system supporting approximate matching".* In Proceedings of the 28th International Conference on Very Large Data Bases, Hong Kong, China, 20–23August2002;pp.1107–1110.

[22] JAGADISH, H.V.; GEHRKE, J.; LABRINIDIS, A.; PAPAKONSTATINOU, Y.; PATEL, J.M.; RAMAKRISHNAM, R.; SHAHABI, C.*" Big Data and Its Technical Challenges*". Communication. ACM 2014, 57, 86–94.

[23] GABRILOVICH, E.*" Feature Genetation for Textual Information Retrieval Using World Knowledge*". Ph.D. Thesis, Palestine Institute of Technology, Haifa, palatine, 2006.

[24] GABRILOVICH, E.; MARKOVICH, S.*" Computing semantic relatedness using wikipedia-based explicit semantic analysis*". In Proceedings of the 20th International Joint Conference on Artifical Intelligence, Hyderabad, India, 6–12 January2007; pp. 1606–1611.

[25] APACHE LUCENE—Welcome to Apache Lucene" https://lucene.apache.org/ "(accessed on 5 March 2017).

[26] HASAN,S.; CURRY, E.*"Thematic Event Processing Dataset*". Available online: https://www.researchgate.net/ publication/263673956_Thematic_event_processing_dataset (accessed on 28 February 2017).

[27] SANCHEZ, L.; GALACHE, J.A.; GUTIERREZ, V.; HERNANDEZ, J.M.; BERNAT, J.; GLUHAK, A.; GARCIA, T. Smart Santander: " The meeting point between Future Internet research and experimentation and the smart cities". In Proceedings of the Future Network & Mobile Summit Conference, Warsaw, Poland, 15–17 June 2011; pp. 1–8.

[28] CURRY, E.; HASAN, S.; O'RIAIN, S. *"Enterprise Energy Management using a Linked Data space for Energy Intelligence*". In Proceedings of the Second IFIP Conference on Sustainable Internet and ICT for Sustainability, Pisa, Italy, 4–5 October 2012.