

الجمهورية العربية السورية  
المعهد العالي للعلوم التطبيقية والتكنولوجيا  
قسم النظم المعلوماتية  
العام الدراسي 2024/2025

## مشروع تخرج

أعد للنجاح في السنة الخامسة

# كشف الأحداث ومراقبتها في وسائل التواصل الاجتماعي اعتماداً على تقنيات التعلم العميق

تقديم

حسن بهجت خضور

إشراف

د. رياض سنبل

8/10/2010

## الخلاصة

يهدف هذا المشروع إلى

## Abstract

Translate your abstract here.

# المحتويات

1	الفصل الأول
1	التعريف بالمشروع
2	1.1- مقدمة
2	2.1- الهدف من المشروع
2	3.1- نطاق المشروع
3	4.1- المتطلبات الوظيفية
4	5.1- المتطلبات غير الوظيفية
5	الفصل الثاني
5	الدراسة المرجعية
6	1.1- مقدمة
6	1.1- مسألة كشف الأحداث في دفق بيانات
7	1.1.1 الإطار العام للمسألة
9	1.1.1 المراحل الرئيسية لسير العمل في نظام لكشف الأحداث
10	1.1.1 تعريف المسألة Problem Definition
10	1.1.1 طرق معالجة المسألة
11	1.1.1 المنهجيات والمقاربات لمعالجة المسألة
13	1.1.1 معايير التقييم وقياس الأداء
13	أ. Normalized Mutual Information (NMI)
13	ب. Adjusted Mutual Information (AMI)
13	ت. B-Cubed
14	1.1.1 مجموعات البيانات المتاحة
16	3.1- مسألة الاستدلال على الموقع الجغرافي من النص
16	1.1.1 الإطار العام للمسألة
17	1.1.1 المنهجيات والمقاربات المتبعة للتعرف على الموقع
18	1.1.1 استخدام نماذج اللغة الكبيرة لاستدلال الموقع المذكور
19	1.1.1 تعريف المسألة Problem Definition

20	1.1.1 مجموعات البيانات المتاحة
20	1.1.1 معايير التقييم وقياس الأداء
21	4.1 - أنظمة مشاهدة
21	Liveuamap.com Live Universal Awareness Map 1.4.1
22	rsoe-edis.org Emergency and Disaster Information Service (EDIS) 2.4.1
23	Globalincidentmap.com Global Incidents Map 3.4.1
25	الفصل الثالث
25	الدراسة النظرية
26	1.1 - مقدمة
26	3.1 - منصات وسائل التواصل الاجتماعي
26	3.1 - العنقدة CLUSTERING
26	1.3.1 العنقدة القائمة على الكثافة Density Based Clustering
27	3.1 - العنقدة ONLINE CLUSTERING
27	3.1 - شبكات المعلومات غير المتجانسة (HIN)
28	2.1 التعرف على الكيانات المسماة NAMED ENTITIES RECOGNITION
28	3.1 GRAPH NEURAL NETWORK
29	الفصل الرابع
29	الدراسة التحليلية
30	1.1 - مقدمة
30	3.1 - مخطط حالات الاستخدام
33	4.1 - السرد النصي لحالات الاستخدام
37	4.1 - العلاقات بين الفاعلين
38	4.1 - مخطط المكونات
39	الفصل الخامس
39	المنهجية المقترحة
40	1.1 - مقدمة
40	2.1 - المقاربة المتبعة لكشف الأحداث
40	3.1 - المقاربة المتبعة لتصنيف الأحداث

40	4.1- منهجية الاستدلال على الموقع الجغرافي
41	الفصل السادس
41	تصميم النظام
42	1.1- مقدمة
42	2.1- تصميم خدمات النظام
42	3.1- مخطط النظام التصميمي
42	4.1- مخطط النشر DEPLOYMENT DIAGRAM
43	الفصل السابع
43	الأدوات المستخدمة
44	1.1- مقدمة
44	2.1- الهدف من المشروع
45	الفصل الثامن
45	تنجيز النظام
46	1.1- مقدمة
46	2.1- تنجيز الخدمات
46	3.1- مخطط النظام
47	الفصل التاسع
47	تحليل ومناقشة النتائج
48	1.1- مقدمة
48	2.1- معايير التقييم
49	الفصل العاشر
49	اختبارات النظام
50	1.1- مقدمة
50	2.1- اختبار خدمات النظام
50	3.1- اختبارات الأداء



## مقدمة عامة

أصبحت وسائل التواصل الاجتماعي اليوم أحد المصادر الرئيسية لنشر المعلومات حول الأحداث اليومية، حيث يقوم المستخدمون بمشاركة الأخبار العاجلة، الكوارث الطبيعية، الجرائم، الحوادث والأحداث الاجتماعية بمجرد وقوعها. ومع النمو السريع في حجم المحتوى المنشور على هذه المنصات، أصبحت هناك حاجة ملحة إلى أنظمة ذكية قادرة على استخراج وتحليل هذه البيانات بشكل آلي لتوفير معلومات دقيقة وتقارير عن الأحداث الجارية.

يعتمد كشف الأحداث على تقنيات معالجة اللغة الطبيعية (NLP) والتعلم العميق لفهم وتحليل المحتوى النصي المنشور على وسائل التواصل الاجتماعي، واستخراج المعلومات ذات الصلة مثل نوع الحدث، وموقعه الجغرافي. كما يتضمن ذلك استخدام تقنيات التعرف على الكيانات المسماة (NER) لاستخلاص الأسماء، الأماكن، والمنظمات المذكورة في النصوص، بالإضافة إلى تقنيات التحليل الجغرافي لتحديد مواقع الأحداث حتى في حال عدم توفر إحداثيات جغرافية مباشرة. تعد هذه العملية ذات أهمية كبيرة في العديد من المجالات، مثل:

- الاستجابة للكوارث والطوارئ: تمكين فرق الإغاثة والجهات الحكومية من الحصول على معلومات دقيقة وفورية حول الكوارث الطبيعية مثل الزلازل والفيضانات والحرائق.
- تحليل الجرائم والأمن العام: مساعدة الأجهزة الأمنية في تتبع الجرائم وحوادث العنف والسرقات فور الإبلاغ عنها عبر وسائل التواصل الاجتماعي.
- رصد الأحداث الاجتماعية والسياسية: تمكين الصحفيين والباحثين من متابعة التغيرات المجتمعية والتطورات السياسية في الزمن الحقيقي.
- إدارة المدن الذكية: تحسين إدارة البنية التحتية والخدمات العامة من خلال مراقبة شكاوى المواطنين وتفاعلهم مع الأحداث المحلية.

يهدف هذا المشروع إلى تقديم نظام متكامل قادر على تحليل تدفقات البيانات الاجتماعية، استخراج المعلومات منها، وعرضها بشكل مرئي يسهل فهمها، مما يساهم في تعزيز الوعي بالأحداث الجارية واتخاذ قرارات مبنية على بيانات دقيقة مرئية.

## الفصل الأول

# التعريف بالمشروع

نبيّن في هذا الفصل هدف المشروع ونطاقه كما نورد المتطلبات الوظيفية وغير الوظيفية.



## 1.1- مقدمة

في وقتنا الحال، أصبحت وسائل التواصل الاجتماعي مصدرنا رئيسياً للأخبار والمعلومات المحلية، فعدت تقدم لنا دفقاً من البيانات من المهم معالجته واستخلاص المعلومات المهمة منه وعرضها بشكل مرئي يساعد الجهات والمنظمات المعنية على اتخاذ القرارات.

## 2.1- الهدف من المشروع

يندرج عملنا في هذا المشروع ضمن سياقين اثنين: (1) الأول نتطرق فيه للمسألة من منظور الذكاء الصناعي، حيث نحاول توظيف تقنياته وأدواته في الوصول إلى مقارنة حل مسألتين مطروحتين هما أ. كشف الأحداث من دفق بيانات وتصنيفها، ب. استنباط الموقع الجغرافي من رسالة نصية، (2) والثاني نتطرق فيه إلى بناء تطبيق برمجي يستفيد من المقاربة المقترحة في تنفيذ التطبيق عملياً مع مراعاة الأسس والمبادئ المتبعة في هندسة البرمجيات وصولاً إلى تطبيق قابل للتوسع وسهل الصيانة.

## 3.1- نطاق المشروع

نطاق هذا المشروع يقتصر على تطوير نظام لاكتشاف الأحداث المحلية باستخدام بيانات وسائل التواصل الاجتماعي وتحليلها باستخدام تقنيات الذكاء الاصطناعي والتعلم العميق. يركز المشروع على تطبيق تقنيات التعلم العميق لتحليل النصوص المستخلصة من وسائل التواصل الاجتماعي، مع الهدف الرئيسي لاكتشاف الأحداث المحلية مثل الحوادث، الأزمات، الجرائم. مكونات المشروع:

1. جمع البيانات: سيتم جمع البيانات من وسائل التواصل الاجتماعي، مثل تويتر، تليغرام، باستخدام واجهات برمجة التطبيقات (APIs) المتاحة لهذه الشبكات وباستخدام الزواحف.
2. تحليل النصوص: استخدام تقنيات التعلم العميق، لتحليل النصوص وتصنيفها إلى أنواع مختلفة من الأحداث (جرائم، حوادث، ...). بالإضافة إلى استخراج الكيانات الأساسية مثل الأماكن والأشخاص المتعلقين بهذه الأحداث باستخدام تقنيات التعرف على الكيانات المسماة (NER).
3. التحليل الجغرافي: التعرف على المواقع الجغرافية من البيانات المستخلصة، سواء عبر البيانات الجغرافية المرفقة (الموسومة جغرافياً) أو من خلال الاستدلال على المواقع. وتطوير آلية لتحويل النصوص الجغرافية إلى إحداثيات جغرافية عبر مرزج جغرافي (Geocoder).

4. النظام التفاعلي: تطوير تطبيق ويب تفاعلي لعرض الأحداث المكتشفة على خرائط تفاعلية في الزمن الحقيقي، مما يسمح للمستخدمين بمراقبة الأحداث وتقييمها على الخريطة. وتقديم تقارير وبيانات إحصائية حول الأحداث المكتشفة.

## 4.1- المتطلبات الوظيفية

نبين هنا المتطلبات الوظيفية التي يجب أن يحققها النظام.

يجب على النظام أن يسمح للمستخدمين بما يلي:

1. استعراض الأحداث حسب المنطقة الجغرافية
2. استعراض الأحداث التي قد مرت (حسب التاريخ)
3. تحديد منطقة جغرافية، لتلقي إشعارات بالأحداث التي تقع بها
4. استخراج مخططات ورسوم توضيحية حول الأحداث وتسلسلها الزمني.

يجب على النظام أن يسمح للمراقب بما يلي:

1. إدارة الأحداث، أي تعديل تصنيفها، حذفها، أهميتها، وإضافة معلومات عنها
- 2.

يجب على النظام ان يسمح لمدير النظام بما يلي:

1. إدارة مصادر البيانات، أي تعديل مصادر سحب البيانات، إضافة مصادر جديدة، حذف مصادر قديمة
- 2.

يجب على النظام أن يقوم بما يلي:

1. جدولة الأحداث بشكل يومي، بحيث يعرض على المستخدم الأحداث اليومية على الخريطة.
2. إرسال إشعارات للمستخدمين عند وقوع أحداث ضمن مناطقهم.
3. استخراج (تجريف Scrape) البيانات من وسائل التواصل الاجتماعي مثل تويتر وتيلغرام والمنصات الإخبارية
4. التخاطب مع APIs وسائل التواصل الاجتماعي وجمع البيانات منها
5. تمرير البيانات في بيئة دقيقة streaming إلى الخدمات الأخرى

## 5.1- المتطلبات غير الوظيفية

نورد هنا المتطلبات غير الوظيفية أي القيود على النظام لكي يتم قبوله. حيث قمنا بتقسيمها إلى عدة محاور وهي متطلبات الأمان ومتطلبات الأداء ومتطلبات خاصة ومتطلبات التشغيل.

### أ. متطلبات الأمان

1. يجب أن يكون الدخول إلى النظام آمناً، أي يجب أن يسمح فقط للمستخدمين المسجلين بالدخول إليه.
2. أن يسمح للمستخدمين القيام بالعمليات وفقاً لما هم مخولين به من صلاحيات فقط.

### ب. متطلبات الأداء

1. يجب أن يستجيب النظام لطلبات المستخدم في غضون زمن محدد لا يتجاوز 2 ثانية لمعظم العمليات.
2. أن يكون النظام قادراً على معالجة دفق البيانات وفق السعة الموضحة بالدراسة الكمية.
3. أن يكون النظام قادراً على استخراج دفق بيانات (data streams) بسعة لا تقل عن X Message /Second

### ت. متطلبات خاصة

1. أن تكون الواجهات باللغة - .

### ث. متطلبات التشغيل

- 1.
2. أن يتم.

## الفصل الثاني

# الدراسة المرجعية

نبيّن في هذا الفصل الأبحاث والأعمال والمنهجيات المشابهة لهذا العمل.

## 1.1- مقدمة

تعد مسألة اكتشاف الأحداث وتصنيفها من وسائل التواصل الاجتماعي من التحديات البحثية الهامة التي تجمع بين عدة مجالات في الذكاء الاصطناعي، مثل معالجة اللغة الطبيعية (NLP)، واستخراج المعلومات، وتحليل البيانات الضخمة. تعتمد هذه المسألة على تحليل دفع من النصوص المنشورة لاستخلاص الأحداث المهمة وتصنيفها وفقاً لنوعها، مثل الجرائم، والحوادث المرورية.

إضافةً إلى ذلك، يُعتبر استنباط الموقع الجغرافي للأحداث من النصوص غير المهيكلة تحدياً إضافياً، حيث يتطلب تقنيات متقدمة لاستخلاص المعلومات الجغرافية وربطها بإحداثيات دقيقة. يهدف هذا القسم إلى تعريف المسألة ومراجعة الأعمال البحثية السابقة في مجال كشف الأحداث وتصنيفها، استخراج الموقع الجغرافي من النصوص، بالإضافة إلى استعراض الأنظمة المشابهة ومجموعات البيانات المتاحة لدعم هذه المهام.

## 1.1- مسألة كشف الأحداث في دفع بيانات

إنَّ التطور السريع لمنصات وسائل التواصل الاجتماعي أدى إلى تضخم في البيانات المنشورة من قبل المستخدمين على الإنترنت. حيث مكنت هذه البيانات الضخمة من دراسة العديد من المشكلات البحثية، وواحد من مواضيعها المهمة هو كشف الأحداث (Li et al., 2022).

وبالانتقال إلى تعريف الحدث، فإن التعريف يختلف قليلاً في الدراسات السابقة حيث يعرفه McMin وآخرون بشكل عام على أنه "شيء مهم يحدث في زمان ومكان محددين"، بينما يعرفه Xie وآخرون على أنه "حدث في العالم الحقيقي، وقع في زمان ومكان معينين، شارك به أو اثر بمجموعة من الأشخاص فحفزهم على التكلم عنه". وبشكل عام يمكننا أن نصفه بما يسمى الـ 4Ws (What, Where, Who & When). بينما يعرف Chao وآخرون الحدث المحلي كما يلي "الحدث المحلي (على سبيل المثال، الاحتجاج، الجريمة، الكوارث الطبيعية) هو نشاط غير عادي ينفجر في منطقة محلية وخلال مدة محددة ويشترك به (أو يتفاعل معه) عدد كبير من المشاركين".

ولطالما تم تناول كشف الأحداث من الوسائط التقليدية في سياق كشف المواضيع وتتبعها Topic Detection and Tracking (TDT)، ومع ذلك فإنَّ اكتشاف الأحداث من وسائل التواصل الاجتماعي يطرح تحدياً جديدة تختلف عن تلك الموجودة في وسائل الإعلام التقليدية. فعلى عكس المقالات الإخبارية، فإن التغريدات تكون مقيدة الطول مما يجعل المعلومات النصية فيها محدودة. كما تشمل كميات كبيرة من الكلمات غير الرسمية والأخطاء الإملائية، وهياكلها غير المنظمة. وتحتوي أيضاً على كميات من الرسائل الوهمية والشائعات (Li et al., 2022).

## 1.1.1 الإطار العام للمسألة

نستطيع النظر إلى مسألة كشف الأحداث من عدة زوايا، استناداً إلى نوع الحدث، يمكننا تصنيف هذه الطرق إلى اكتشاف حدث غير محدد وحدث محدد Unspecified Vs. Specified. وفي مجال ال TDT، يصنف كشف الأحداث إلى فئتين (1) كشف الأحداث الجديدة (NED) New Event Detection و (2) كشف الأحداث الرجعية Retrospective Event Detection (RED). Event Detection ففي حالة NED يكون هدفنا هو اكتشاف الأحداث الجديدة في دفع بيانات في الزمن الحقيقي، بينما في حالة RED فإننا نركز على التعرف واكتشاف الأحداث في بيانات تاريخية (Li et al., 2022).

يمكننا الآن مناقشة كل نوع على حدى، فنبداً إذن بالتصنيف القائم على النوع:

### ♣ كشف الحدث المحدد Specified Event Detection

في هذا النوع من الأحداث يكون لدينا علم مسبق بمكان وزمان وقوع الحدث. أي عندما يكون الحدث الاجتماعي معروفاً أو مخططاً له بالفعل، فإن معالجة البيانات المتعلقة بالمعلومات المعروفة (مثل الموقع والوقت والكلمات الرئيسية والمستخدمين) لاستخراج وصف الحدث تسمى اكتشاف الأحداث المحددة (SED). ويعالج في هذا النوع، المعلومات المحددة مسبقاً والميزات التي من المتوقع أن تظهر في البيانات لتمثيل حدث. تعمل هذه المعلومات المحددة مسبقاً كبذرة لسياق الحدث الفعلي. يمكن التعبير عن المعلومات المتعلقة بالحدث كلياً أو جزئياً لجمع البيانات وتحليلها (Li et al., 2022).

### ♣ كشف الحدث غير المحدد Unspecified Event Detection

في هذا النوع من الأحداث، لا علم لنا بمكان وزمان وقوع الحدث ولا بطبيعته، فعندها يكون مصدرنا الوحيد لتحديد هو مراقبة الخصائص الزمانية والكلمات المنتشرة. أي عندما تكون المعلومات الاجتماعية حول حدث ما غير معروفة، يعتمد النهج الأساسي ل UED على تحليل الأنماط الزمنية لتدفق البيانات من خلال مراقبة التدفق لتحديد الكلمات الرئيسية والمفاهيم المتكررة ذات الصلة بتسليط الضوء على الأحداث (Li et al., 2022).

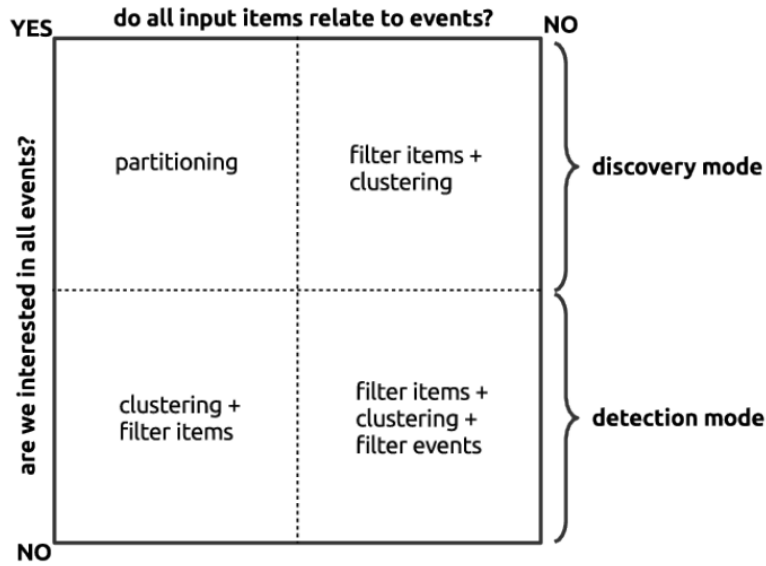
أما بالنسبة للتصنيف الثاني:

### ♣ كشف الأحداث الجديدة والرجعية New Event Detection & Retrospective Event Detection

اعتماد على المهمة ونوع الحدث، يمكننا أيضاً تصنيف الحدث إلى نوعين NED و RED. ونظراً لأن تقنيات NED تتضمن المراقبة المستمرة لتدفق البيانات لكشف الأحداث الجديدة في الزمن الفعلي، فإنه يكون مناسباً بطبيعته لكشف أحداث العالم الحقيقي غير المعروفة والأخبار العاجلة. كما يمكن استخدام تقنيات ال NED من أجل كشف الأحداث المحددة، على الرغم من أن معظم الدراسات تركت على الأحداث غير المحددة. وعندما تتضمن المهمة المرادة أحداثاً محددة أو معلومات محددة

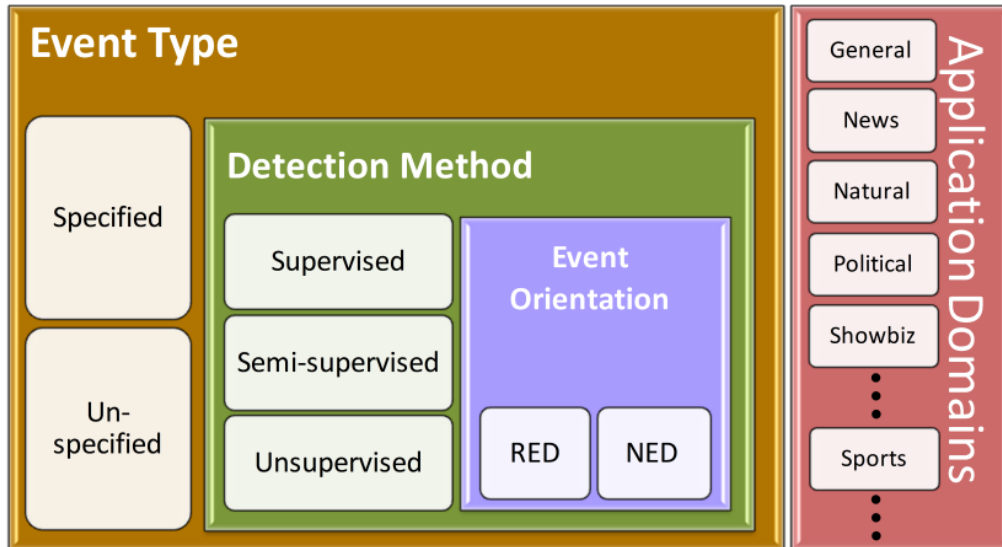
حول الحدث (عل سبيل المثال، منظمة محددة، شخص أو مكان ما) يمكن الاستفادة من هذه المعلومات لمكاملتها مع تقنيات ال NED من خلال تقنيات الفلترة أو التصنيف.

كما ويوضح الشكل أدناه بعض نطاقات المعالجة المهمة (تجزئة، فلترة، عنقدة) وذلك حسب نمط المسألة. أي إذا كنا نهتم بجميع الأحداث أو إذا كانت جميع الرسائل تنتمي إلى أحداث (Schinas et al., 2018).



الشكل 1: نطاقات معالجة كشف الأحداث حسب نمطها (Schinas et al., 2018).

ويختصر الشكل أدناه مفردات المسألة ويصنفها بشكل هرمي يوضحها.



الشكل 2: تصنيف هرمي للأحداث (Saeed et al., 2019).

### 1.1.1 المراحل الرئيسية لسير العمل في نظام لكشف الأحداث

يقسم Li وآخرون في بحثهم مراحل عمل نظام كشف الأحداث لثلاثة مراحل كما هو موضح في الشكل أدناه، هي (1) مرحلة المعالجة المسبقة والتي تتم على دفع البيانات الذي يرد إلى النظام، (2) مرحلة كشف الأحداث، (3) مرحلة مابعد كشف الحدث.

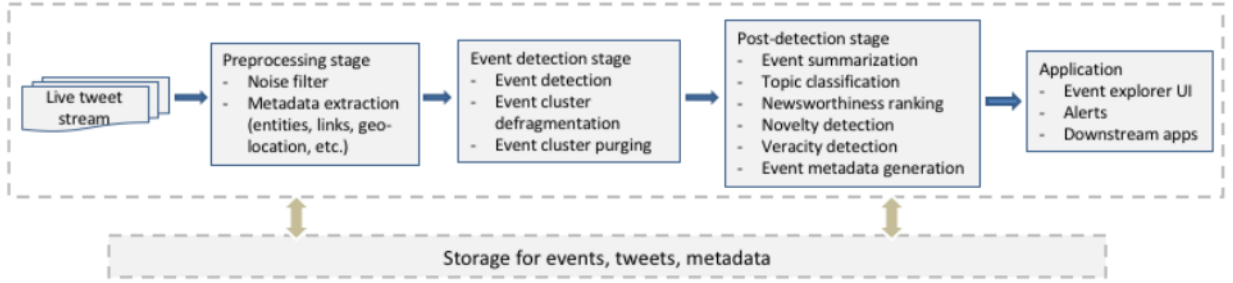


Figure 1 : The typical workflow for a NED system for unspecified events (Li et al., 2022).

ومهمة كل مرحلة هي كما يلي:

- ♣ مرحلة المعالجة المسبقة Preprocessing Stage، تحوي هذه المرحلة عدة مكونات مثل مكون فلتر الضجيج الذي يقوم بإزالة الرسائل غير المهمة، الإعلانات وماشابه. وتحتوي على مكونات لاستخراج الكيانات، المعلومات الجغرافية.
- ♣ مرحلة كشف الحدث Event Detection Stage، اعتماداً على نوع الحدث المراد كشفه (Specified, Unspecified, RED, NED) قد تختلف تقنية الكشف الفعلية (على سبيل المثال، الطرق المعتمدة على العقدة، المعتمدة على المصطلحات، ...). وتهدف مكونات Event Cluster Defragmentation إلى دمج مجموعات الأحداث ذات الصلة، و Purging Event إلى إزالة الأحداث القديمة التي لم يعد هناك حاجة إليها.
- ♣ مرحلة مابعد الكشف Post-Detection Stage، حسب التطبيق الذي نهتم به، قد نحتاج إلى بعض المكونات مثل تلخيص الأحداث Event Summarization، تحديد الموضوع Topic Detection، كشف الشائعات.



### 1.1.1 تعريف المسألة Problem Definition

أ- تعريف: دفع البيانات الاجتماعي A Social Stream

نعرف دفع البيانات  $\mathcal{S}$  على أنه تسلسل مرتب زمنياً ومستمر لكتل الرسائل الاجتماعية Social Messages Blocks، ممثلة كـ  $\mathcal{S} = M_0, \dots, M_{i-1}, M_i, \dots$  حيث كل كتلة رسائل  $M_i$  تحوي كل الرسائل التي وصلت خلال الشريحة الزمنية  $[t_i, t_{i+1}]$  أي  $M_i$  هي مجموعة من الرسائل الفردية، أي  $M_i = \{m_j; 1 \leq j \leq |M_i|\}$ ، حيث كل رسالة  $m_j$  تحوي نص الرسالة بالإضافة إلى المرسل وزمن الإرسال (Abagissa et al., 2024).

ب- تعريف: الحدث An Event

نرمز له بـ  $e$ ، وهو مجموعة من الرسائل الاجتماعية المترابطة المرتبطة بحدث في العالم الحقيقي. ونفترض أن كل رسالة مرتبط بحدث اجتماعي واحد على الأكثر. أي  $e = \{m_j; 1 \leq j \leq |e|\}$  (Abagissa et al., 2024).

ت- تعريف: خوارزمية كشف الأحداث Event Detection Algorithm

من أجل كتلة رسائل  $M_i$ ، يتم تدريب خوارزمية كشف الأحداث Social ED Algorithm لتعلم نموذج  $\mathcal{F}$ ، بحيث  $\mathcal{F}(M_i; \theta) = E_i$  حيث  $E_i$  هي مجموعة الأحداث التي تظهر في الكتلة  $M_i$ . أي  $E_i = \{e_j; 1 \leq j \leq |E_i|\}$ . و  $\theta$  هو موسط كدخل للنموذج (Abagissa et al., 2024).

### 1.1.1 طرق معالجة المسألة

هنالك تنوع كبير في منهجيات كشف الأحداث، ونستطيع تصنيف هذه المنهجيات إلى أربع فئات اعتماداً على مبدأ عملهم:

♣ مرتكزة على السمة Feature-Pivot، تعتمد هذه الطرائق على كشف أنماط غير طبيعية في بعض سمات النص (على سبيل المثال، بعض العبارات المتفجرة Bursty Phrases)، أي عندما يقع حدث ما فإن بعض الكلمات والعبارات سيتم تداولها بشكل كثير مما يجعل توزيعها غير طبيعي بالنسبة لتوزيعها تاريخياً، وبناءً على ذلك يتم الكشف.

♣ مرتكزة على المستند Document-Pivot، تعتمد هذه الطرائق على تحويل المستند (message & its meta data) إلى شعاع، ومن ثم عنقدة الأشعة وتجميعها بناءً على التشابه. وفي عدد من الدراسات الحديثة، تم تعزيز هذه المنهجية باستخدام الشبكات العصبية البيانية (GNNs)، حيث يتم الاستفادة من شبكات المعلومات غير المتجانسة لتحويل بيانات وسائل التواصل الاجتماعي من شكلها غير المهيكل إلى شكل مهيكل ذو علاقات دلالية، ومن ثم الاستفادة من البيان الناتج ومن قدرات GNNs في تمثيل العلاقات غير الخطية والتفاعلات المعقدة بين الكيانات. لتحسين وزيادة فعالية تمثيل اشعة التضمين، على سبيل المثال، وظفت أبحاث مثل (KPGNN - Cao et al., 2021)، (HCRC - Guo et al., 2024) و (QSGNN - Ren et al., 2022)\* تقنيات GNNs.

♣ مرتكزة على الموضوع Topic-Pivot، تضمن هذه الطرائق استخدام نماذج إحصائية لكشف الحدث على اعتباره متحول ضمني في المستند. (Schinas et al., 2018).

وفي بعض الأبحاث لجاء الباحثون إلى طريقة قائمة على البيان Graph Based، حيث يتم الاستفادة من الهيكلية الدلالية للرسائل ضمن وسائل التواصل الاجتماعي حيث يتم تحويل مجموعة الرسائل غير المهيكلة Unstructured Messages إلى رسائل مهيكلة ذات دلالة Semantic Structured Data (وذلك باستخدام الكيانات المسماة، hashtags، عمليات الذكر) وذلك بإنشاء بيان، وبعد إنشاء هذا البيان تتحول المسألة حينها على مسألة تجزئة البيان Graph Partitioning.

### 1.1.1 المنهجيات والمقاربات لمعالجة المسألة

يوضح الجدول أدناه الأبحاث ذات الصلة والمقاربات المتبعة لكشف الأحداث في وسائل التواصل الاجتماعي.

#	Paper (Name)	Techniques	Supervision	Event Type	Detection Task
1	Petrović et al., 2010	Online Clustering & LSH	Unsupervised	U	NED
2	Sakaki et al., 2010	SVM	Supervised	S	RED
3	Long et al., 2011	Hierarchical Clustering	Unsupervised	U	NED
4	Cordeiro, 2012	Wavelet Analysis & LDA	Unsupervised	U	NED
5	McCreadie et al., 2013	K-Mean Clustering & LSH		U	NED
6	Corney et al., 2014	Online Clustering		S	NED
7	Alsaedi et al., 2015	Online Clustering	Unsupervised	U	NED
8	Patil et al., 2015	NN Classification	Supervised	S	-
9	Guille & Favre, 2015	Mention, Anomaly Detection		U	NED
10	Zhang et al., 2016	Spatial-temporal Clustering	Unsupervised	U	NED
11	Alsaedi et al., 2016	Temporal, Spatial, Textual Features, Online Clustering	Unsupervised	U	NED
12	Chen et al., 2017	NN, Online Clustering		U	NED
13	Zhang et al., 2017	Spatial-Temporal Clustering	Unsupervised	U	NED
14	Edouard, 2018	Domain Vocabulary and Knowledge Based		S	NED
15	Comito et al., 2019	Text & Temporal Feature Clustering	Unsupervised	U	NED
16	Hossny et al., 2019	Keyword Volume Approach, SVM	Supervised	S	NED
17	Fedoryszak et al., 2019	Cluster Chain, Graph	Unsupervised	U	NED
18	Nguyen et al., 2019	Entity Clustering		U	NED
19	Saeed et al., 2019	Graph, Clustering	Unsupervised	U	NED
20	Morabia et al, 2019 (SEDTWik)	Segments, Hierarchical Clustering	Unsupervised	U	NED

21	Liu et al., 2020 (EventX)	Community Detection	Unsupervised	U	NED
22	Cao et al., 2020 (IED KCN)	Knowledge Consolidation Network (KCN)	Supervised	-	NED
23	Li and Zhang, 2021	Semantic Term, GCN, Clustering		U	NED
24	Ghaemi et al., 20221	Spatial-Temporal Clustering, VDBSCAN	Unsupervised	U	NED
25	Cao et al., 2021 (KPGNN)	GNN, Clustering	Supervised	-	NED
26	George et al., 2021	Spatial-Temporal Clustering	Unsupervised	U	NED
27	Mojiri et al., 2021	Burst Keywords, Time Series Analysis	Unsupervised	U	NED
28	Yamani et al., 2021	Knowledge Based, MLM	Supervised	-	-
29	Peng et al., 2022 (FineEvent)	GNNs	Supervised	-	-
30	Hettiarachchi et al., 2022 (Embed2Detect)	NN, hierarchical clustering	Unsupervised	U	NED
31	Afyouni Khan et al., 2022	Spatial-Temporal Clustering, Classification	-	U	NED
32	Ren et al., 2022 (QSGNN)	GNNs, Online Clustering	Supervised	-	-
33	Bok et al., 2023	Graph, Online Clustering	Unsupervised	U	NED
34	Cao et al., 2024 (HISEvent)	GNNs, Graph Clustering	Unsupervised	U	
35	Yue t al., 2023	MLM	Supervised	-	-
36	Pradhan et al., 2024 (EDTBert)	Graph Clustering, PLMs			
37	Guo et al., 2024 (HCRC)	GNNs, Incremental Clustering	Unsupervised	U	NED
38	Ray1 et al., 2024	Incremental Clustering	Unsupervised	U	NED
39	RPLMSED	PLMs	Supervised	S	NED
40	Abagissa et al., 2024 (DistilBert – GNN)	GNNs, Clustering		U	NED
41	Yang et al., 2024 (ADPSEM)	Structural Entropy Minimization	Unsupervised	-	-
42	Yu et al., 2024 (HyperSED)	Hyperbolic Embedding, Graph Partitioning	Unsupervised	U	
43	Kovalchuk et al., 2025 ( SemConvTree)	Semantic Convolutional Tree	Unsupervised	U	NED

### 1.1.1 معايير التقييم وقياس الأداء

#### أ. Normalized Mutual Information (NMI)

المعلومات المتبادل المعيرة، يقيس مقدار المعلومات المشتركة بين المجموعات الحقيقية (Ground Truth) والمجموعات المكتشفة (Predicted Clusters). يتم تقيسه باستخدام الإنتروبي ليصبح المقياس بين 0 و 1. وهو مفيد لتقييم جودة العنقدة مع أخذ احتمالية الاتفاق العشوائي بعين الاعتبار.

$$NMI(U, V) = \frac{I(U; V)}{H(U) + H(V)}$$

حيث  $I(U; V)$  هي المعلومات المتبادلة بين المجموعتين  $H(U)$  و  $H(V)$  هي الإنتروبي لكل من المجموعات الحقيقية والمتوقعة.

#### ب. Adjusted Mutual Information (AMI)

مقياس معدل من NMI يأخذ بالحسبان مقدار المعلومات المتبادلة المتوقعة (Expected Mutual Information) عند التوزيع العشوائي للتسميات. بالتالي، يوفر مقياسًا أكثر دقة في الحالات التي يكون فيها الاتفاق العشوائي مرتفعًا.

$$AMI(U, V) = \frac{I(U; V) - \mathbb{E}[T(U; V)]}{\max(H(U), H(V)) - \mathbb{E}[T(U; V)]}$$

#### ت. B-Cubed

مقاييس B-Cubed Precision و B-Cubed Recall تستخدم لتقييم نتائج العنقدة (Clustering) بحيث تأخذ بعين الاعتبار كل عنصر ومقارنته مع عناصر أخرى في نفس المجموعة. أي بعكس المقاييس الأخرى التي تقيم التوزيع الكامل للمجموعات دفعة واحدة، يقوم B-Cubed بتقييم كل عنصر على حدة، ويحسب دقة واسترجاعه بناءً على الانتماءات الفعلية والمتوقعة.

$$BCubed\ Precision = \frac{1}{N} \sum_{e_i \in E} \frac{|Correct\ Labels(e_i) \cap Cluster(e_i)|}{|Cluster(e_i)|}$$

$$BCubed\ Recall = \frac{1}{N} \sum_{e_i \in E} \frac{|Correct\ Labels(e_i) \cap Cluster(e_i)|}{|Correct\ Labels(e_i)|}$$

### 1.1.1 مجموعات البيانات المتاحة

يوضح الجدول أدناه مجموعات البيانات المتاحة والمستخدمة في معظم المراجع لتدريب النماذج لهذه المسألة وقياس أدائها.

	Dataset	Size	Events	Language	Paper
1	Kawarith	~9K tweet	7	Arabic	(Alharbi & Lee, 2021)
2	Events 2012	~68K tweet	503	English	(McMinn et al., 2013)
3	Maven	~10K tweet	164	English	(Wang et al., 2020)
4	Events 2018	~64K twee	257	French	-
5	FA Cup	~670K tweet		English	-
6	CrisisLex	~60K tweet	6	English	(Olteanu et al., 2014)
7	Few Events	~8K tweet	64	English	
8	KBP	~85K tweet	100	English	(Deng et al., 2020)
9	FloDust	~20K tweet	3	Arabic	(Hamoui et al., 2021)
10	CrisisMMD	~18K tweet	7	English	-

جدول 1: مجموعات البيانات المتاحة لمسألة كشف الأحداث.

كما ويوضح الجدول أدناه مجموعات البيانات التي تم استخدامها في الأبحاث، حيث يلاحظ أن معظم الأبحاث يعتمد على مجموعة بيانات خاصة بالبحث (غير متوفرة).

	Paper   Dataset	Events 2012	Events 2018	Maven	Few Events	CrisisLex	Kawarith	FA Cup	Other	Private
1	Alsaedi et al. , 2015									Y
2	Zhang et al. , 2017									Y
3	Hossny et al. , 2019									Y
4	Saeed et al. , 2019							Y	Y	
5	Fedoryszak et al. ,2019	Y								
6	Morabia et al. ,2019	Y								
7	Zhang et al.,2021									Y
8	Hettiarachchi et al. ,2021							Y		
9	Ghaemi et al. ,2021									Y
10	Mojiri et al. ,2021							Y	Y	
11	Yamani et al. ,2021								Y	
12	Cao et al. ,2021	Y		Y						
13	Peng et al. ,2021								Y	
14	Afyouni et al. ,2022									Y
15	Ren et al. ,2022	Y	Y							
16	Kolajo et al. ,2022									Y
17	Peng et al. ,2022									Y
18	Bok et al. ,2023									Y
19	Cao et al. ,2023	Y	Y							
20	Ren et al. ,2023 a	Y	Y				Y			
21	Ren et al. ,2023 b	Y	Y			Y				

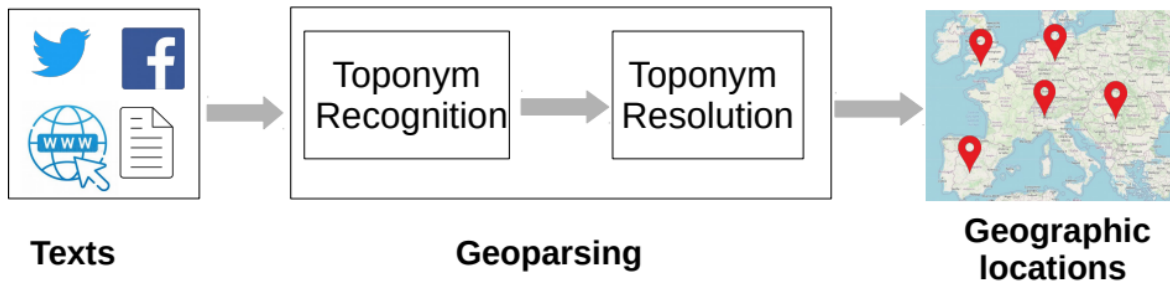
	Paper   Dataset	Events 2012	Events 2018	Maven	Few Events	CrisisLex	Kawarith	FA Cup	Other	Private
22	Guoa et al. ,2023			Y						
23	Yue et al. ,2023			Y	Y					
24	Yang et al. ,2024	Y	Y							
25	Qiu et al. ,2024	Y				Y	Y			
26	Ray et al. ,2024	Y	Y							
27	Yu et al. ,2024	Y	Y				Y		Y	
28	Abagissa et al. ,2024	Y								
29	Pradhana et al. ,2024							Y	Y	
30	Ma et al. ,2024	Y	Y		Y	Y	Y			
31	Li et al. ,2024	Y	Y				Y			
32	Kovalchuk et al. ,2024 a									Y
33	Kovalchuk et al. ,2024 b	Y							Y	
34	Kargupta et al. ,2024									Y
35	Murthy et al. ,2025									Y

جدول 2: مجموعات البيانات المستخدمة في الأبحاث المدروسة.

### 3.1- مسألة الاستدلال على الموقع الجغرافي من النص

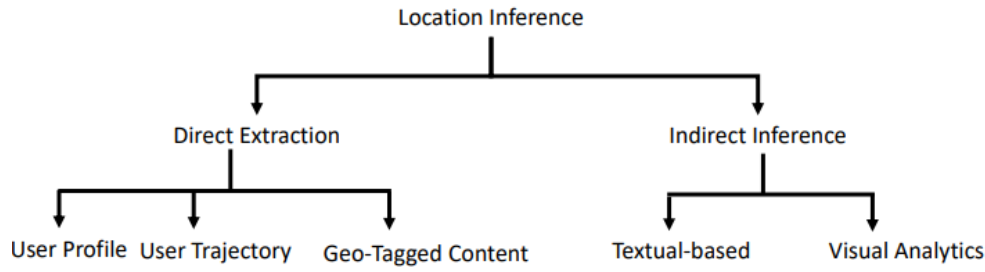
#### 1.1.1 الإطار العام للمسألة

يعرف الاستدلال على الموقع من النصوص الفردية -ويسمى أيضاً (GeoParsing)- بعملية تقدير الإحداثيات الجغرافية للنص بناءً على محتواها أو بياناتها الوصفية أو كليهما. ونظراً لأن نسبة ضئيلة فقط من النصوص في مواقع التواصل الاجتماعي تحمل علامات جغرافية صريحة geotag in metadata فعلى سبيل المثال، في منصة تويتر فقط 3% من التغريدات تحمل علامات جغرافية صريحة، ولذلك يعتمد الباحثون على المعلومات المستخرجة من اللغة الطبيعية - مثل أسماء الأماكن المذكورة في نص التغريدة - بالإضافة إلى المعلومات السياقية مثل الوسوم المعلومات الزمنية وبيانات المستخدم. ويتكون نظام الاستدلال على الموقع الجغرافي من عمليتين هما (1) التعرف على الموقع المذكور Location Mention Recognition (LMR)، و (2) تحديد الموقع وفك غموضه Location Mention Disambiguation and Detection (LMD). وبالنسبة لعملية التعرف على الموقع المذكور LMR والتي تسمى في بعض المراجع أيضاً بـ Toponym Recognition وهو حالة خاصة من مسألة التعرف على الكيانات المسماة Named Entities Recognition (NER) حيث يتم فيها تحديد الكلمات التي تدل على الموقع واستخراجها. أما في عملية تحديد الموقع وفك غموضه والتي أيضاً تسمى في بعض المراجع بـ Toponym Resolution، ففيها يتم تحويل المواقع المستخرجة من المرحلة السابقة إلى إحداثيات جغرافية أو معرفات في قاموس جغرافي Gazetteer، وذلك بعد فك الغموض الذي يحدث نتيجة وجود عدد من العناوين المتشابهة (مثل مدينة بانباس في محافظة طرطوس، ومدينة بانباس في محافظة القنيطرة) التي تحل من خلال أخذ المعلومات السياقية بالاعتبار أو من خلال بعض التجريبيات. ويوضح الشكل أدناه آلية العمل تلك بشكل عام.



رسم توضيحي 1: مخطط العمل العام ضمن نظام استدلال الموقع (Afyouni et al., 2022).

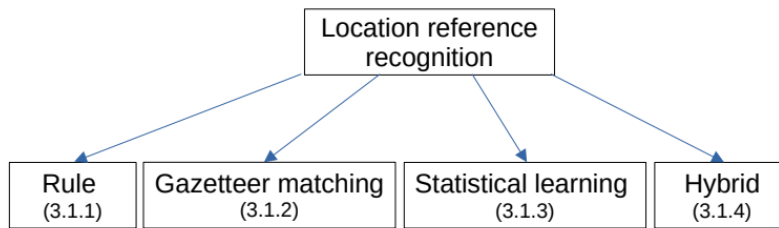
كما ويصنف الباحثون مصادر المعلومات المستخدمة للاستدلال على الموقع، إلى فرعين، الأول استخراج مباشر يتم عن طريق استخراج المعلومات المكانية الموسومة من ملف المستخدم الشخصي أو عن طريق المعلومات المكانية الموسومة مع النص المنشور حيث تقدم بعض وسائل التواصل مثل تويتر وسوم مكانية في حال فعل المستخدم الـ GPS، والفرع الثاني هو غير مباشر ويكون إما عن طريق الاستدلال من خلال النص المنشور (وهو المناسب في حالتنا) أو عن طريق تحليل المحتوى المرئي (مثل الصور) وهذا لن نتطرق له في حالتنا. والشكل أدناه يوضح هذه الأنماط.



الشكل 3: أنماط لاستدلال على الموقع الجغرافي (Afyouni et al., 2022).

### 1.1.1 المنهجيات والمقاربات المتبعة للتعرف على الموقع

تصنف المقاربات المتبعة للتعرف على الموقع ضمن أربع أنماط هي (1) معتمدة على القواعد، (2) معتمدة على المطابقة من Gazeteer، (3) معتمدة على التعلم، (4) طرق هجينة بين الثلاث السابقة.



الشكل 4: تصنيف طرائق التعرف على الموقع.

❖ مقاربات قائمة على القواعد Rule-Based Approaches: في هذا النمط من الأنماط يتم معالجة المسألة من خلال الأنماط المنتظمة Regular Expressions، وذلك لأن الموقع الذي يذكر في النص غالباً ما يحتوي على سمات تميزه لذلك يتم الاعتماد على قواعد (سواء كانت موضوعة من قبل خبراء، أم تم تعلمها) لاستخراج النص الذي يشير إلى موقع، وهذه الطرق قليلة الاستخدام نظراً لأن النص المكتوب في وسائل التواصل الاجتماعي يحوي على ميزات تجعل تطبيق هذا النوع غير مجدي، ولكن تبقى هذه الطرق مستخدمة لتحسين أداء المقاربات الأخرى عند استخدامها معها.



- ❖ مقاربات مطابقة القاموس الجغرافي Gazetteer Matching Approaches: في المقاربات القائمة على المطابقة من قاموس جغرافي، يعتمد الباحثون في هذه المقاربة على أخذ الـ n-gram للنص، ومن ثم لكل سلسلة يقومون بمحاولة لمطابقتها مع القاموس الجغرافي، وبعد ذلك يتم فلتر النتائج وفك غموضها بناء على مجموعة تجريبية heuristics.
- ❖ مقاربات التعلم Statistical Learning Approaches: في هذه الطرائق يتم الاعتماد على تعلم الآلة Machine Learning وعلى مدونة منمطة، لتعلم نموذج يقوم بتصنيف المفردات في النص. وفي معظم الأبحاث يقوم الباحثون بالاعتماد على نماذج التعرف على الكيانات المسماة NER ومن ثم ضبطها وتدريبها على هذه المسألة Fine-Tuning، وفي أبحاث أخرى كما (Chen et al., Bobadilla et al., Cadorel et al.) يقوم الباحثون بالتعامل مع المسألة بشكل منفصل عن NER وتدريب نماذجهم لهذه المسألة خصيصاً.
- ❖ المقاربات الهجينة Hybrid Approaches: في هذا النوع يلجأ الباحثون إلى استخدام المقاربات السابقة مع بعضها البعض بحيث يتم الاستفادة من ميزات كل مقاربة لتحسين الحل المقترح.

### 1.1.1 استخدام نماذج اللغة الكبيرة لاستدلال الموقع المذكور

لقد عزز ظهور نماذج اللغة الضخمة (LLMs) بشكل كبير قدرة الآلة على فهم استفسارات المستخدم المعقدة، وعزز فعالية معالجة اللغة وفهمها لها، مما أفاد العديد من المجالات، بما في ذلك مجال العلوم الجغرافية المكانية. وقد بحثت العديد من الدراسات في إمكانيات وقيود نماذج اللغة الضخمة في هذا المجال (Ji and Gao 2023, Mooney et al. 2023, Tao and Xu 2023, Xie et al. 2023, Yin et al. 2023, Hochmair et al. 2024). فعلى سبيل المثال، درس Xie وآخرون حدود تطبيق نماذج الذكاء الاصطناعي في السياقات الجغرافية المكانية، ودعوا إلى نماذج ذات أساس جغرافي مُصممة خصيصاً لها. حيث يسلط هذا البحث الضوء على الاختلاف بين أنواع البيانات التقليدية، مثل نصوص اللغة الطبيعية والفيديو، التي صُممت من أجلها نماذج الذكاء الاصطناعي، والبيانات الجغرافية المكانية، مُشيراً إلى ضرورة الابتكار في هذا المجال. وقيم Ji و Gao مدى كفاءة نماذج اللغة الضخمة في تفسير وتمثيل المفاهيم المكانية المذكورة في النصوص. وقرن Hochmair وآخرون أداء نماذج ChatGPT-4 و Bard و Claude-2 و Copilot في مهام جغرافية مكانية مُختلفة، مثل فك الغموض المكاني، ومفاهيم نظم المعلومات الجغرافية، ورسم الخرائط. كما واستكشف Juhasz وآخرون إمكانية استخدام GPT 3.5-turbo لإثراء Open Street Map (OSM) من خلال اقتراح أفضل الوسوم لكل طريق في OSM بناءً على أوصاف مشتقة من صور Mapillary. وبعض الأبحاث أيضاً حول تطبيق نماذج اللغة الضخمة (LLM) في التحليل الجغرافي، على سبيل المثال، قدّم Li وآخرون نموذج لغة مكاني GeoLM، وهو مدرب مسبقاً باستخدام بيانات من مصادر مثل ويكيبيديا و Wikidata و OpenStreetMap. يمكن تكييف GeoLM لدعم مهام متعددة في مراكز المدن، مثل التعرف على أسماء المواقع الجغرافية، والتحليل، واستخراج العلاقات، وكتابة الكيانات الجغرافية. كما واستكشف Mai وآخرون إمكانية الاستفادة من نماذج اللغة الضخمة، في مجال الذكاء الاصطناعي الجغرافي. وأثبتوا أن نماذج اللغة الضخمة تتفوق على النماذج المخصصة لمهام محددة في مهام الدلالات

الجغرافية المكانية، بما في ذلك التعرف على أسماء المواقع الجغرافية والتعرف على وصف الموقع. وقام أيضاً Hu وآخرون بدراسة جدوى نماذج GPT المختلفة، لاستخراج أوصاف المواقع من وسائل التواصل الاجتماعي أثناء الكوارث. وقارنوا هذه النماذج بأدوات التعرف على الكيانات المسماة (NER) التقليدية ونموذج BERT المعدل. استخدمت الدراسة مجموعة بيانات من التغريدات المنمطة من إعصار هارفي. وتُظهر النتائج أن نماذج GPT الموجهة بالمعرفة الجغرافية تفوقت بشكل ملحوظ على أدوات NER ونموذج BERT المعدل في التعرف على أوصاف المواقع الكاملة والفئات المرتبطة بها. كما وقيم Hu وآخرون أداء نماذج اللغة الضخمة المخففة Lightweight في مهمة التعرف على الموقع المذكور وفك الغموض.

### 1.1.1 تعريف المسألة Problem Definition

كما ذكرنا آنفاً أنّ المسألة تتكون من جزئين التعرف على الموقع المذكور (LMR)، وتحديد الموقع وفك الغموض (LMD)، ونستطيع صياغة المسألة لهما بشكل رياضي كما يلي:

١) تصاغ مسألة التعرف على الموقع المذكور كما يلي

ليكن لدينا نص  $t$ ، يهدف نظام LMR إلى تحديد جميع المواقع المذكورة في النص  $L_t = \{l_i; i \in [1, n_t]\}$  حيث  $n_t$  هي العدد الكلي للمواقع المذكورة في النص  $t$ . وكل  $l_i$  قد تغطي أكثر من مفردة في النص لذلك يتم استخدام نموذج الـ BILOU، حيث B توسم بداية الموقع و I توسم أننا ضمن ذكر موقع و L توسم نهاية ذكر الموقع و O أن المفردة خارج الرمز بينما ترمز U إلى أن ذكر الموقع يغطي مفردة واحدة فقط (Suwaileh et al., 2023).

٢) تصاغ مسألة تحديد الموقع وفك غموضه كما يلي

بعد أن يتم التعرف على جميع المواقع المذكورة في النص  $L_t = \{l_i; i \in [1, n_t]\}$ ، يتم هنا ربط كل ذكر موقع  $l_i$  بتمثيل جغرافي دقيق وفريد، وهو يكون إما عبارة عن إحداثيات ( $latitude, longitude$ ) أو معرف في قاعدة بيانات جغرافية. تُعرّف هذه المسألة رياضياً كتابع  $LMD : L \rightarrow G$ ، حيث  $L$  هو مجموعة المواقع المذكورة المستخرجة من النص  $t$ ،  $G$  هي مجموعة الكيانات الجغرافية الممكنة (على سبيل المثال، Gazeeter)، حيث تعطي لكل مدخل جغرافي إحداثياته ( $latitude, longitude$ ). حيث يهدف التابع LMD إلى إيجاد المطابقة الأمثل لكل  $l_i$  بحيث يُسند إلى  $l_i$  الكيان الجغرافي  $g \in G$  الذي يمثل بدقة المقصود في السياق النصي (Suwaileh et al., 2023).

### 1.1.1 مجموعات البيانات المتاحة

يوضح الجدول أدناه مجموعات البيانات المتاحة والمستخدمه في معظم المراجع لتدريب النماذج لهذه المسألة وقياس أدائها.

	Dataset	Size	Region	Language	Paper
1	IDRISI-R (Gold)	~20K En, ~4.6K Ar	Global (41 disaster events)	Ar, En	(Suwaileh et al., 2023)
2	IDRISI-R (Silver)	~57K En, ~1.2M Ar	Global (41 disaster events)	Ar, En	(Suwaileh et al., 2023)
3	GeoCorpora	~6,648 tweets	USA	En	
4	GeoCoV19	~524M tweets	Global (COVID-19 related)	Multilingual	(Qazi et al., 2020)
5	UTGEO	~670K tweets	USA	Multilingual	
6	TweeLoc	~325K tweets	Rome, Netherlands	Un Specified	

جدول 3: مجموعات البيانات المتاحة لمسألة التعرف على الموقع المذكور.

### 1.1.1 معايير التقييم وقياس الأداء

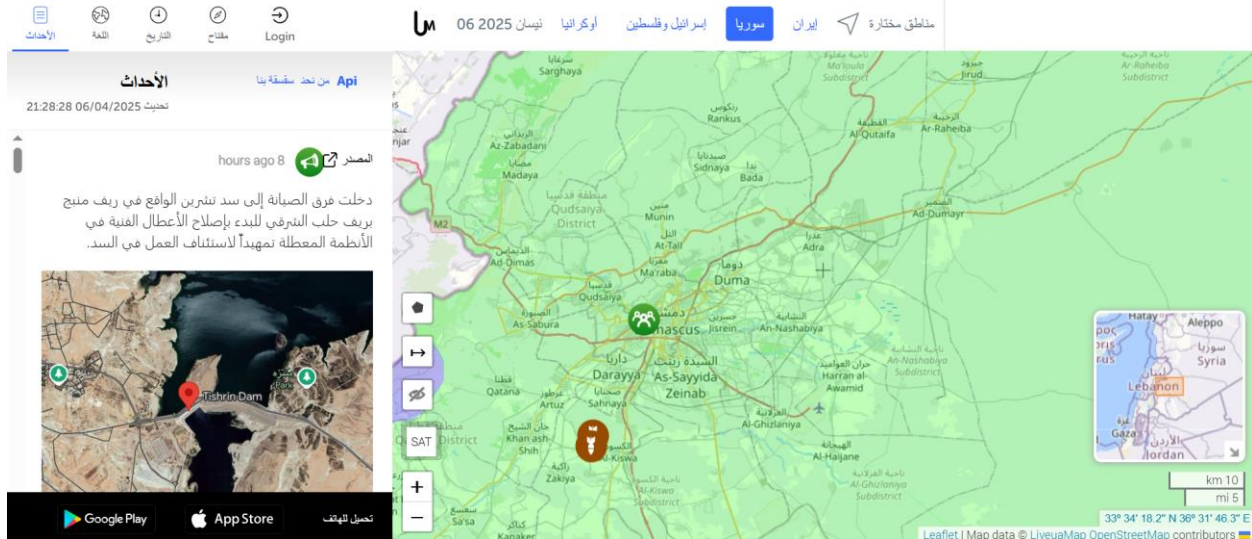
من أجل تقييم النتائج لهذه المسألة وقياس أداء النماذج يوجد العديد من معايير التقييم، التي تختلف حسب الهدف من النموذج فمن أجل النماذج التي تهدف إلى التنبؤ بـ (المنطقة، المدينة، ...) فيتم في هذه الحالة استخدام المعايير التقليدية كالصحة، الدقة والاستدكار (Accuracy, Precision, Recall)، أي يتم التعامل معاً كمسألة تصنيف طبيعية. (أي هي مناسبة في حالة التعرف على الموقع المذكور (LMR))، أما بالنسبة للنماذج التي تتنبأ بالإحداثيات الجغرافية للرسالة، ففي هذه الحالة يتم استخدام معايير التقييم المستخدمة لمسائل الانحدار متوسط خطأ المسافة Mean Error Distance، الصحة في الكيلومتر Accuracy@KM (أي بتكميم المسألة وتحويلها إلى تصنيف ضمن خلايا)، مسافة هافيرسين (Haversin Distance) وهي صيغة مسافة بين الإحداثيات الجغرافية ويستخدم لقياس المسافة على سطح الأرض، وصيغته هي كما هو موضح أدناه.

## 4.1- أنظمة مشابهة

يوجد عدد من المنصات التي تعالج المسألة المطروحة سواء بشكل شبه كامل أو جزء منها، على مستويات مختلفة، أي يوجد بعض المنصات التي تعالج المسألة على مستوى عالمي وأخرى على مستوى محلي.

### Liveuamap.com Live Universal Awareness Map 1.4.1

خريطة الوعي العالمي الحي، هو موقع إلكتروني تفاعلي يعرض خريطة عالمية تحتوي على أخبار وأحداث جارية يتم تحديد مواقعها جغرافيًا. المنصة تُستخدم لتتبع الأحداث السياسية، العسكرية، الكوارث الطبيعية، وأخبار الأمن والنزاعات في الوقت الحقيقي، حيث يسمح للمستخدم بتحديد نوع الأحداث التي يريد الاطلاع عليها (كوارث طبيعية، نزاعات، ...) ومن ثم يعرض عليها بشكل حي هذه الأحداث ويتم تحديث المعلومات بالزمن شبه الحقيقي. حيث يعد مصدرًا قيمًا للمعلومات، ويوفر وصولاً مفتوحًا عبر الإنترنت إلى أرشيف زمني كامل للمعلومات، مما يتيح للمشاهدين البحث في الأحداث الماضية والاتجاهات التاريخية



#### ♣ كيف يعمل الموقع؟

- يعتمد على مصادر متعددة لجمع البيانات مثل وسائل التواصل الاجتماعي، وكالات الأنباء، والمصادر المحلية.
- تستخدم المنصة خوارزميات آلية ومحررين بشريين للتحقق من صحة الأخبار وربطها بالموقع الجغرافي المناسب.
- يتم تصنيف الأحداث إلى أنواع متعددة (مثل: نزاع، سياسة، طقس، صحة، كوارث، إلخ) باستخدام رموز مختلفة.
- يمكن للمستخدمين استعراض الأخبار حسب الموقع الجغرافي، أو حسب التصنيف الزمني أو الموضوعي.

#### ♣ كيف يجمع البيانات؟

- يعتمد على زواحف زكية AI Crawlers لتجميع البيانات وانتقاء الأخبار الجديرة بالذكر.

- يتم جمع البيانات من عدة مصادر من بينها وسائل التواصل الاجتماعي.
- بالإضافة إلى خوارزميات والنماذج، يعتمد الموقع على خبراء بشريين.

#### 2.4.1 rsoe-edis.org Emergency and Disaster Information Service (EDIS)

موقع RSOE EDIS (نظام معلومات الطوارئ والكوارث)، هو خدمة تدار من قبل الجمعية الوطنية المجرية للإغاثة اللاسلكية والاتصالات، ويهدف إلى تتبع وتوثيق وتحليل الكوارث والحوادث الطارئة حول العالم. ويهدف إلى إبقاء العامة، والمنظمات الإنسانية، والباحثين، والحكومات على اطلاع دائم بالأحداث الطارئة، حيث يعرض بيانات حيّة تم جمعها من وكالات الأنباء، وسائل الإعلام، المراسد، ... إلخ، ويقوم بتحليلها وعرض الجدير بالذكر منها.



صورة 1: الصفحة الرئيسية للموقع rsoe-edis.org

#### ♣ كيف يعمل الموقع؟

- يقوم بجمع البيانات من جهات رسمية ومنظمات إنسانية ووسائل إعلام موثوقة.
- يجمع الموقع البيانات من مجموعة كبيرة من المصادر مثل:
  - وكالات حكومية
  - وسائل الإعلام الرسمية
  - منظمات دولية
  - شبكات المتطوعين
- يعرض الأحداث في الوقت الحقيقي عبر خريطة تفاعلية عالمية.
- يوفر اشعارات للمستخدمين وتنبهات.

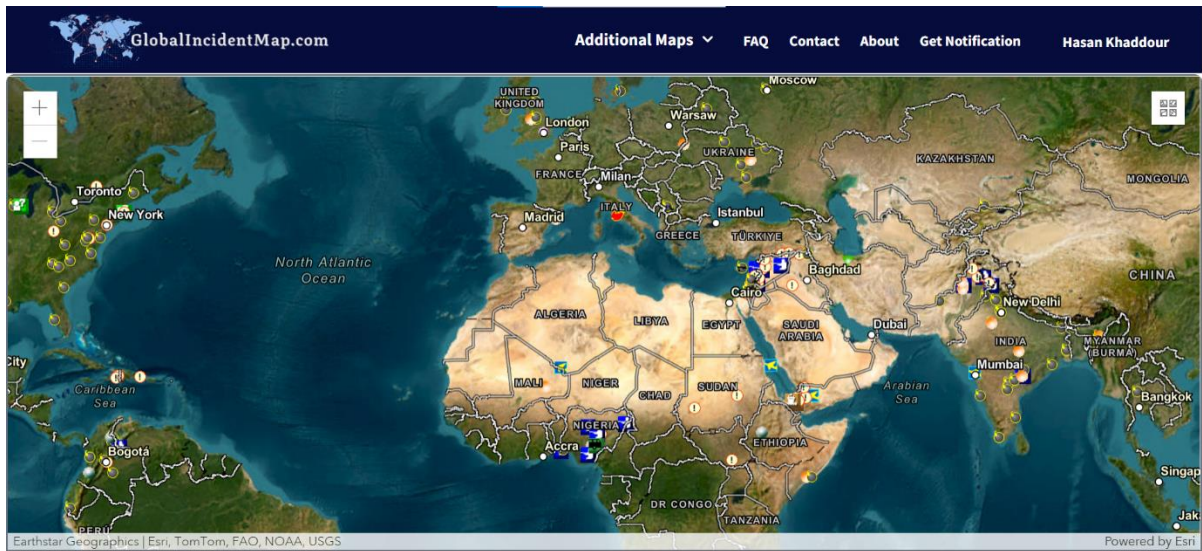
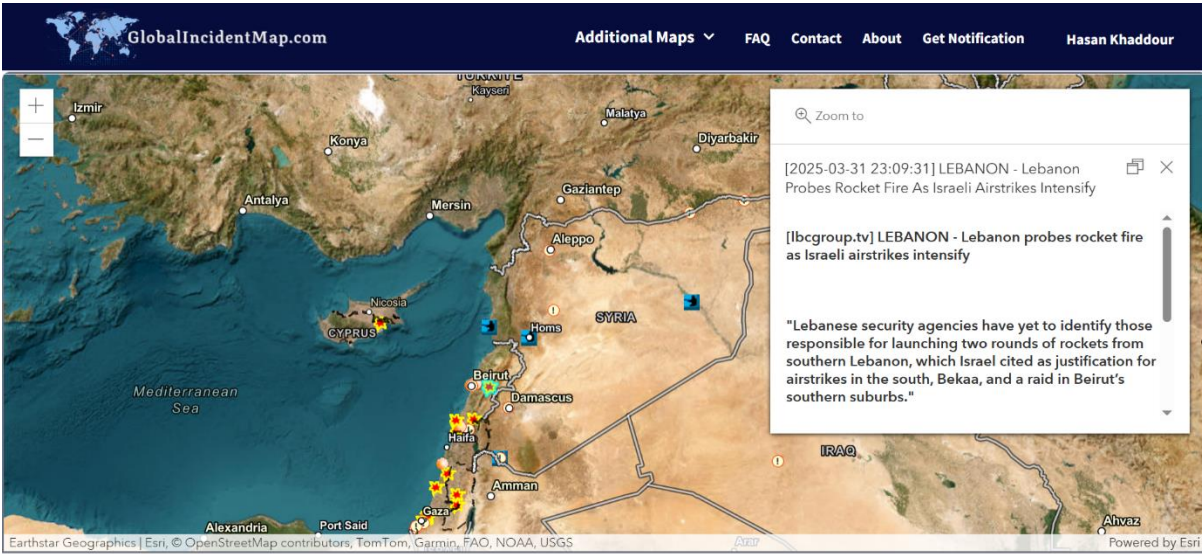


• يوفر تقارير تفصيلية حول أنواع مختلفة من الكوارث مثل:

- الكوارث الطبيعية (زلازل، فيضانات...)
- التهديدات المتعلقة بالبنية التحتية
- الحوادث البيولوجية
- الأزمات الاجتماعية

### 3.4.1 Globalincidentmap.com Global Incidents Map

هو موقع إلكتروني يعرض خريطة تفاعلية لحوادث وتهديدات أمنية من مختلف أنحاء العالم. يركز على التقارير المتعلقة التهديدات الأمنية، الحوادث البيولوجية، اختطاف الطائرات، وأحداث أخرى تمس الأمن العام.



صورة 2: الصفحة الرئيسية للموقع globalincidentmap.com

#### ♣ كيف يعمل الموقع؟

- يعرض الأحداث على خريطة زمنية-مكانية تفاعلية تُحدث بشكل مستمر.
- يتم تصنيف الحوادث في فئات متعددة (مثل: تهديدات أمنية، إطلاق نار، حوادث طيران، حرائق كبيرة، كوارث طبيعية، إلخ).
- يمكن للمستخدمين تصفح الحوادث حسب الفئة، التاريخ، أو الموقع الجغرافي.
- يوجد عدد أنماط للموقع أحدها هو الحوادث، بينما يوجد أيضاً العديد من الأنماط
- منها الكوارث الطبيعية، أزمات الغذاء، الأزمات السيبرانية إلخ.

#### ♣ كيف يجمع البيانات؟

- يعتمد على مزيج من المصادر الإخبارية الرسمية والمصادر المفتوحة مثل تقارير الشرطة، وكالات الأنباء العالمية، والمصادر عبر الإنترنت.
- يتم فلتره وتصنيف البيانات من خلال أنظمة ذكية بمساعدة محررين بشريين وخوارزميات تحليل نصوص.
- بعض الخرائط يتم تحديثها يدوياً، وبعضها يتم تغذيته عبر تدفقات بيانات شبه فورية.

#### ♣ ما الفائدة من هذا النظام؟

- أداة قوية لمتابعة التهديدات العالمية والأحداث الأمنية لحظة بلحظة.
- تُستخدم من قبل الصحفيين، مراكز الدراسات الأمنية، وفرق إدارة الأزمات.
- تُوفر تنبيهات في الوقت الحقيقي للمستخدمين المسجلين.

## الفصل الثالث

# الدراسة النظرية

يوضح هذا الفصل بعض المفاهيم المستخدمة في هذا العمل.



## 1.1- مقدمة

### 3.1- منصات وسائل التواصل الاجتماعي

#### 3.1- العنقدة Clustering

العنقدة، وهي عملية من تقنيات تعلم الآلة غير الخاضعة للإشراف unsupervised، حيث تقوم على تجزئة مجموعة من البيانات (العناصر) إلى مجموعات جزئية --أي عناقيد-- بحيث تكون العناصر الموجودة في العنقود الواحد متشابهة مع بعضها البعض، وتختلف عن العناصر الموجودة في المجموعات الأخرى. وتصنف طرق العنقدة بشكل عام في أربع مجموعات: (1) طرائق التجزئة Partitioning Based، حيث يتم تحديد عدد العناقيد مسبقاً، أو (2) القائمة على الشبكة Grid Based، حيث يتم تقسيم فضاء العناصر إلى عدد محدد مسبقاً من الخلايا، أو (3) هرمية Hierarchical، حيث يتم تنظيم البيانات في مستويات متعددة، أو (4) قائمة على الكثافة Density Based، حيث يتم أخذ الكثافة بالاعتبار (Igartua et al., 2020).

#### 1.3.1 العنقدة القائمة على الكثافة Density Based Clustering

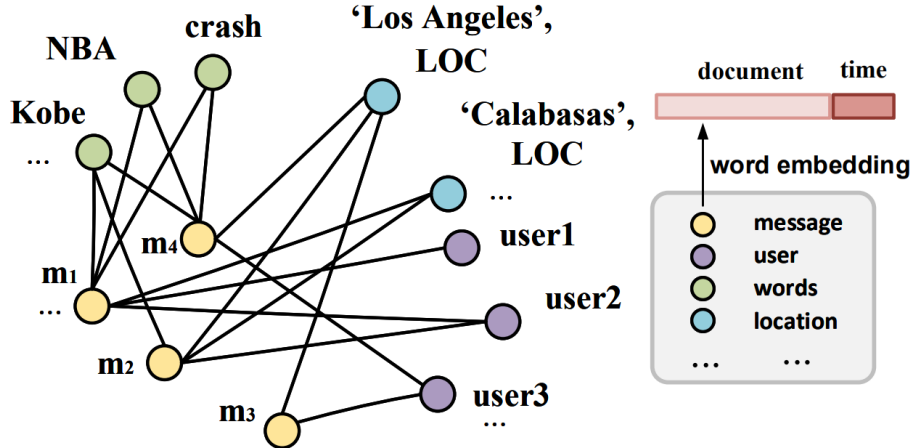
تنظم خوارزميات العنقدة القائمة على الكثافة البيانات في المناطق (العناقيد) حيث تكون العناصر كثيفة والتي تفصل عن بعضها بمناطق ذات كثافة عناصر منخفضة، والتي تعتبر ضجيج. وهذه الخوارزميات قادرة على (1) اكتشاف مجموعات من الأشكال الاعتبائية arbitrary، (2) التعامل مع المناطق المفرغة sparse (التي تعتبر مناطق ضجيج) و (3) العمل دون معرفة عدد العناقيد مسبقاً (Igartua et al., 2020). ومن بين المقترحات المختلفة في الأدبيات يوجد DBSCAN (التجميع المكاني القائم على الكثافة للتطبيقات ذات الضجيج Density-Based Spatial Clustering of Applications with Noise). DBSCAN هي خوارزمية شائعة للعنقدة القائمة على الكثافة، تجمع النقاط المتراسة مع بعضها البعض، مع تمييز النقاط في المناطق منخفضة الكثافة كقيم شاذة. حيث وبخلاف K-means، لا تتطلب DBSCAN تحديد عدد العناقيد مسبقاً، ويمكنها تحديد العناقيد ذات الأشكال المختلفة. يعتمد أدائها على معاملي:  $\epsilon$  (نصف قطر الجوار) و MinPts (الحد الأدنى لعدد النقاط اللازمة لتكوين منطقة كثيفة). تتميز DBSCAN بفعالية خاصة في قواعد البيانات المكانية والتحليل الجغرافي المكاني (Ester et al., 1996).

### 3.1- العنقدة Online Clustering

تشير العنقدة الحية، والمعروفة أيضًا بالعنقدة التزايدية (Incremental Clustering)، أو العنقدة التدفقية (Streaming Clustering)، إلى تقنيات العنقدة التي تحدّث العناقيد ديناميكيًا مع وصول بيانات جديدة، دون الحاجة إلى إعادة معالجة مجموعة البيانات بأكملها. ويكتسب هذا أهمية خاصة في الحالات التي تتضمن بيانات واسعة النطاق أو آنية، مثل شبكات الاستشعار، والتوصيات عبر الإنترنت، وتدفقات وسائل التواصل الاجتماعي. تهدف خوارزميات مثل خوارزمية K-means التزايدية النسخ الحية من DBSCAN إلى التكيف بكفاءة مع انحراف المفاهيم وضمان قابلية التوسع مع ذاكرة محدودة (Aggarwal et al., 2003).

### 3.1- شبكات المعلومات غير المتجانسة (HIN)

شبكات المعلومات غير المتجانسة (HIN) Heterogeneous Information Networks، وهي حالة خاصة من البيان Graph، حيث يحتوي هذا على أكثر من نوع من العقد والوصلات. أي كما يلي،  $G = (V, E, \mathcal{N}, T)$  حيث  $V$  و  $E$  هما مجموعات العقد والوصلات و  $\mathcal{N}$  و  $T$  هما مجموعات أنواع العقد والوصلات، حيث لكل عقدة نوع من  $\mathcal{N}$  و لكل وصلة نوع من  $T$ ، فعلى سبيل المثال في حالتنا، تكون الرسالة  $m$  وهي عقدة (من نمط رسالة) مرتبطة مع عقد الأماكن المذكورة في نصها (وهذه العقد تكون من نوع أماكن)، وتكون أيضاً هذه الرسالة مرتبطة مع عقد الـ Hashtags المذكورة فيها (وهي عقد من نوع Hashtag)، وبذلك وبالمرور على كل الرسائل الواردة في هذه الشريحة المنية، نكون قد شكلنا هذه الشبكة. ويوضح الشكل أدناه مثالاً عن ذلك.



رسم توضيحي 2: مثال على شبكة معلّزمات غير متجانسة.

## 2.1- التعرف على الكيانات المسماة Named Entities Recognition

يعرف الكيان المسمى Named Entity بأنه كائن واقعي يملك تعريفاً مناسباً، ويمكن الإشارة له باسمه المناسب. حيث يمكن أن تكون الكيانات المسماة مكاناً أو شخصاً أو منظمة أو عنصراً أو كياناً جغرافياً أو وقتاً. وكمثال عن الكيانات المسماة في "المعهد العالي" هو كيان مسمى من نوع مؤسسة. وتتم عملية التعرف على الكيانات المسماة باستخدام أدوات وتقنيات معالجة اللغات الطبيعية، حيث تقوم بتصنيف مفردات النص المدخل إلى فئات محددة مسبقاً تمثل الكيانات المسماة.

## 3.1- Graph Neural Network

## الفصل الرابع

# الدراسة التحليلية

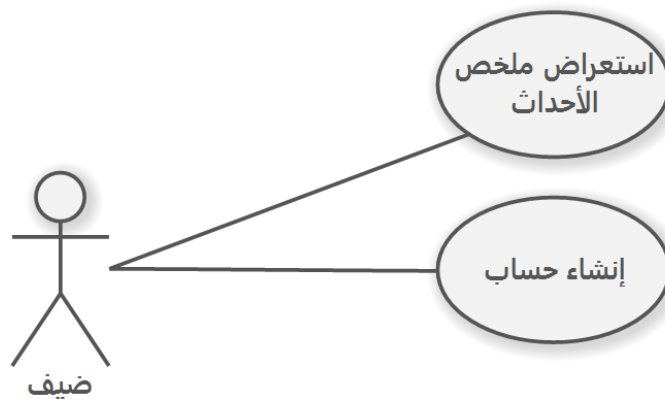
يقدم هذا الفصل تحليلاً للمتطلبات التي أوردناها في الفصل الأول.

## 1.1- مقدمة

نقدم في هذا الفصل دراسة تحليلية للمسألة المطروحة، ونبين حالات الاستخدام ووصفها السردى وبعض مخططات UML الداعمة.

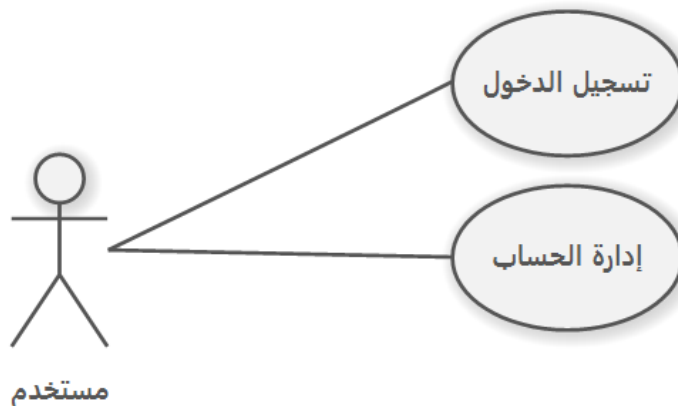
### 3.1- مخطط حالات الاستخدام

أ- حالات الاستخدام الخاصة بالضيف



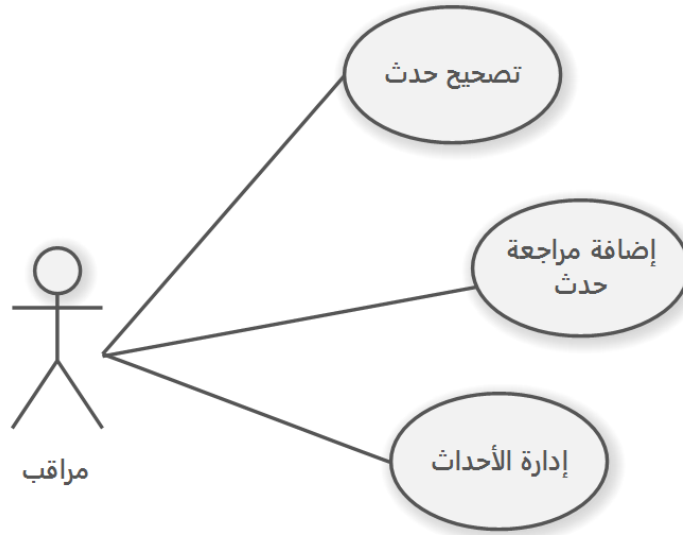
رسم توضيحي 3: مخطط حالات الاستخدام للضيف.

ب- حالات الاستخدام الخاصة بالمستخدم



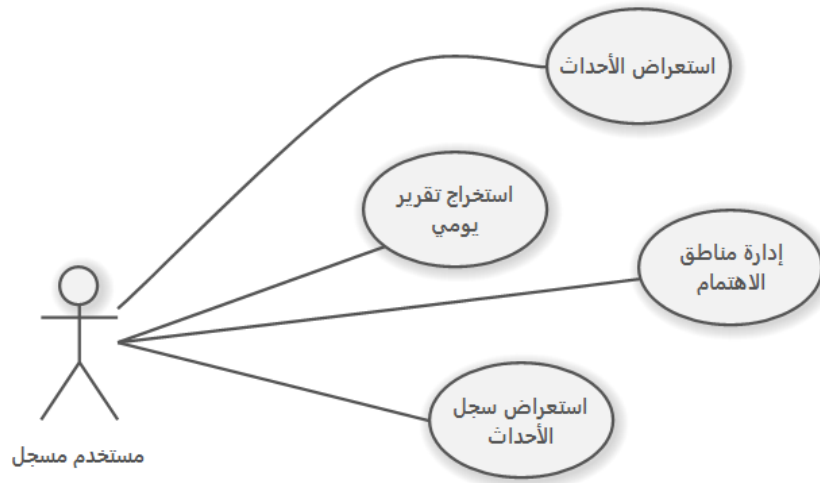
رسم توضيحي 4: مخطط حالات الاستخدام للمستخدم.

ت- حالات الاستخدام الخاصة بالمراقب



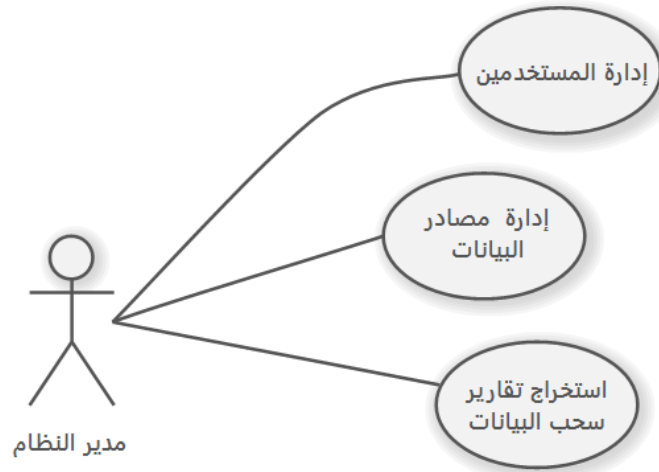
رسم توضيحي 5: مخطط حالات الاستخدام للمراقب.

ث- حالات الاستخدام الخاصة بالمستخدم المسجل



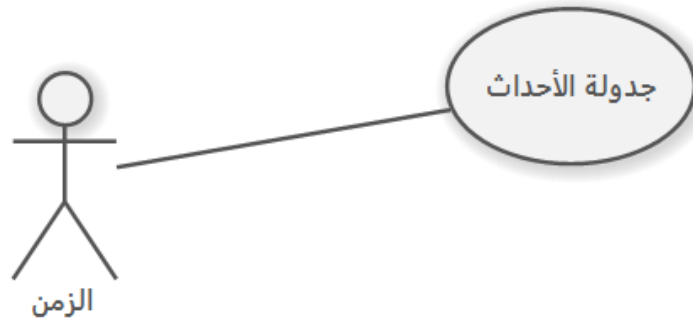
رسم توضيحي 6: مخطط حالات الاستخدام للمستخدم المسجل.

ج. حالات الاستخدام الخاصة بمدير النظام



رسم توضيحي 7: مخطط حالات الاستخدام لمدير النظام.

ح. حالات الاستخدام التي يفعلها الزمن



رسم توضيحي 8: مخطط حالات الاستخدام التي يفعلها الزمن.

## 4.1- السرد النصي لحالات الاستخدام

نوضح هنا السرد النصي لحالات الاستخدام الرئيسية وهما جدول الأحداث، إدارة مناطق الاهتمام، استعراض الأحداث.

أ. حالة الاستخدام جدول الأحداث

اسم حالة الاستخدام: جدول الأحداث	
الفاعلون الأوليون	الزمن
الشروط المسبقة	• لا يوجد.
الشروط اللاحقة	• تمت جدول الأحداث التي حصلت اليوم، أي إزالتها من قائمة الأحداث المعروضة وإضافتها إلى سجل الأحداث التاريخية.
الوصف	يتم جدول الأحداث التي حصلت اليوم وإضافتها إلى سجل الأحداث، وذلك في ساعة محددة مسبقاً.

سير الأحداث

### السيناريو الرئيسي الناجح – Main Success scenario

جدول 4: السيناريو الناجح لحالة الاستخدام جدول الأحداث

النظام	الزمن
1. يبدأ عملية جدول الحدث.	
2. يستخرج قائمة الأحداث اليومية	
3. يضيف قائمة الأحداث اليومية إلى سجل الأحداث	
4. يستخرج التقرير اليومي للأحداث	
5. يرسل إشعاراً إلى مدير النظام بنجاح الجدولة، ويرفق معه التقرير المستخرج.	
6. ينهي النظام عملية الجدولة.	



## المسارات البديلة

A1: في المرحلة رقم 2-أو 3 أو 4، في حال فشل إحدى عمليات الاستخراج، فإنّ النظام يرسل إشعاراً إلى مدير النظام بحدوث خطأ ويرفق العملية التي تسببت بالخطأ، ومن ثم يعيد المحاولة بعد الانتظار لزمان محدد، ويتابع من المرحلة رقم 1. وفي حال تكرار المحاولة أكثر من ثلاث مرات، فعندئذٍ ينتقل إلى حالة الخطأ E1.

## المسارات الخاطئة

E1: في حال تكرار السيناريو البديل أكثر من ثلاث مرات، فعندئذٍ ينهي النظام عملية الجدولة.

ب. حالة الاستخدام استعراض الأحداث

اسم حالة الاستخدام: استعراض الأحداث	
المستخدم المسجل	الفاعلون الأوليون
• المستخدم مسجل بالنظام، ومعروف لدى النظام، أي تم تنفيذ حالة الاستخدام "تسجيل الدخول" بنجاح.	الشروط المسبقة
• لا يوجد.	الشروط اللاحقة
يقوم المستخدم المسجل باستعراض الأحداث على الخريطة أو بشكل قائمة، والاطلاع على ملخصها، واختيار أحدها والاطلاع عليه بشكل مفصل.	الوصف

## سير الأحداث

### السيناريو الرئيسي الناجح – Main Success scenario

جدول 5: السيناريو الناجح لحالة الاستخدام استعراض الأحداث

المستخدم المسجل	النظام
1. يقوم المستخدم ببدء عملية الاستعراض	
	2. يعيد قائمة بأنماط العرض الممكنة (خريطة، قائمة، تقرير report).
3. يختار نمط العرض المراد.	

4. يعيد مجموعة الأحداث اليومية وفق نمط العرض المختار.	
5. يستعرض الأحداث، وبعد ذلك ينهي عملية الاستعراض	

#### المسارات البديلة

A1: في المرحلة رقم 5، إذا أراد المستخدم استعراض حدث ما بشكل تفصيلي فيتابع كما يلي ويكمل من المرحلة رقم 5:

1. يختار الحدث الذي يريده، ويرسل معرفه	
2. يرسل معلومات الحدث بشكل تفصيلي (ملخص، تقييم، مراجعات المراقب، ...)	

#### المسارات الخاطئة

لا يوجد.

ت. حالة الاستخدام إدارة مناطق الاهتمام

اسم حالة الاستخدام: إدارة مناطق الاهتمام	
المستخدم المسجل	الفاعلون الأوليون
• المستخدم مسجل بالنظام، ومعروف لدى النظام، أي تم تنفيذ حالة الاستخدام "تسجيل الدخول" بنجاح.	الشروط المسبقة
• لا يوجد.	الشروط اللاحقة
يقوم المستخدم المسجل باستعراض المناطق التي يهتم بتلقي اشعارات بالأحداث التي تقع بها، إضافة منطقة جديدة ، أو إزالة منطقة اهتمام.	الوصف

#### سير الأحداث

السيناريو الرئيسي الناجح – Main Success scenario

جدول 6: السيناريو الناجح لحالة الاستخدام إدارة مناطق الاهتمام

النظام	المستخدم المسجل
1. يقوم المستخدم ببدء عملية الاستعراض	
2. يعيد النظام قائمة المناطق التي يهتم بها المستخدم.	
3. ينهي المستخدم عملية الاستعراض.	

#### المسارات البديلة

A1: في المرحلة رقم 3، إذا أراد المستخدم إضافة منطقة جديدة، فيتابع كما يلي ومن ثم يكمل من المرحلة رقم 3:

1. يختار بدء عملية إضافة منطقة اهتمام	
2. يعيد خريطة المناطق، والنطاقات الممكنة.	
3. يختار المنطقة والنطاق (الإحداثيات، ونصف القطر)	
4. يضيف النظام المنطق الجديدة إلى قائمة المناطق.	
5. ينهي عملية إضافة المنطقة.	

A1: في المرحلة رقم 3، إذا أراد المستخدم إزالة منطقة جديدة، فيتابع كما يلي ومن ثم يكمل من المرحلة رقم 3:

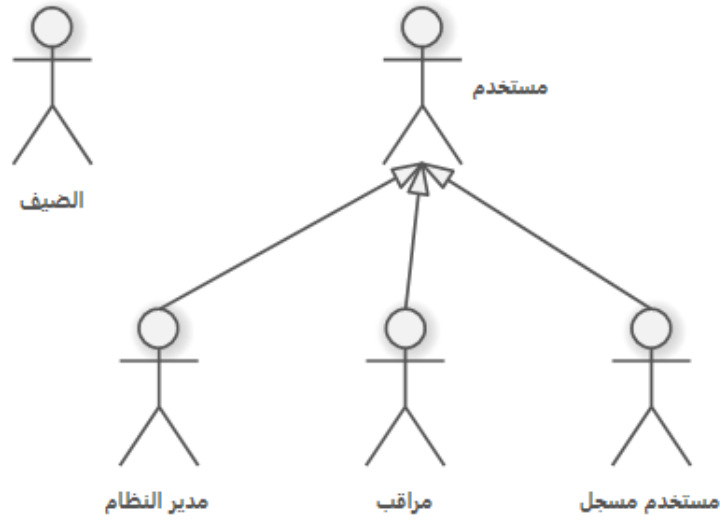
1. يختار بدء عملية إزالة منطقة اهتمام	
2. يعيد قائمة المناطق المهتم بها	
3. يختار المنطقة التي يريد إزالتها	
4. يزيل النظام المنطقة من قائمة المناطق.	
5. ينهي عملية الإزالة المنطقة.	

#### المسارات الخاطئة

لا يوجد.

## 4.1- العلاقات بين الفاعلين

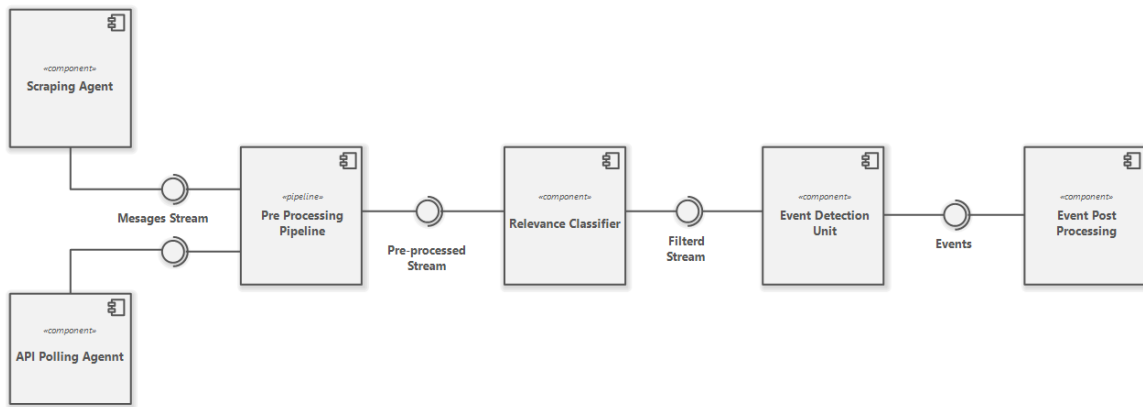
يوضح المخطط أدناه العلاقات بين الفاعلين في هذا النظام.



رسم توضيحي 9: العلاقات بين الفاعلين.

## 4.1- مخطط المكونات

يساعد المخطط أدناه في فهم مكونات النظام، حيث يتكون من عدة مكونات، وهي مكونان من أجل سحب البيانات الأول يقوم بسحب البيانات من خلال الاستقصاء النشط لـ API المقدمة من وسائل التواصل (تويتر وتيلغرام) والثاني يقوم بسحبها من خلال تقليد سلوك مستخدم حقيقي، ومن ثم يقدمان دفق البيانات للمكون التالي الذي يقوم بتطبيق المعالجات المسبقة على دفق الرسائل ومن ثم يأتي دور المكون الثالث الذي يقوم بفلتر الدفق بالاستعانة بالمصنف المدرب على التعرف على الرسائل ذات الصلة، ويمررها بدوره لمكون كشف الأحداث، الذي يصنفها إلى عناقيد ويمررها إلى المكون الأخير الذي يقوم بمعالجة لاحقة عليها. وهذا يوضح مكونات النظام بشكل عام ومجرد بعيداً عن التفاصيل التقنية.



الشكل 5: مخطط مكونات النظام.

## الفصل الخامس

# المنهجية المقترحة

يعرض هذا الفصل منهجية العمل المقترحة لتحقيق النظام بناءً على المنهجيات التي عرضناها في الفصل الثاني.

## 1.1- مقدمة

تعتب

## 2.1- المقاربة المتبعة لكشف الأحداث

يبل

## 3.1- المقاربة المتبعة لتصنيف الأحداث

يهدف

## 4.1- منهجية الاستدلال على الموقع الجغرافي

يهدف هذا الجزء كما تشير تسميته إلى

## الفصل السادس

# تصميم النظام

يوضح هذا الفصل بعض المفاهيم المستخدمة في هذا العمل.



## 1.1- مقدمة

تعتبر

## 2.1- تصميم خدمات النظام

نطلب التقيد

## 3.1- مخطط النظام التصميمي

يهدف

## 4.1- مخطط النشر Deployment Diagram

يهدف

## الفصل السابع

# الأدوات المستخدمة

نعرض في هذا الفصل الأدوات وأطر العمل المستخدمة في تنفيذ النظام.

## 1.1- مقدمة

.

## 2.1- الهدف من المشروع

## الفصل الثامن

# تنجيز النظام

يوضح هذا الفصل بعض المفاهيم المستخدمة في هذا العمل.

## 1.1- مقدمة

تعتبر .

## 2.1- تنجيز الخدمات

نطلب التقيد

## 3.1- مخطط النظام

يهدف

.

## الفصل التاسع

# تحليل ومناقشة النتائج

يوضح هذا الفصل بعض المفاهيم المستخدمة في هذا العمل.

## 1.1- مقدمة

تعتبر الصفحات

## 2.1- معايير التقييم

ن

## الفصل العاشر

# اختبارات النظام

يوضح هذا الفصل بعض المفاهيم المستخدمة في هذا العمل.



## 1.1- مقدمة

تعتبر الـ.

## 2.1- اختبار خدمات النظام

نطلب

## 3.1- اختبارات الأداء

ي

## الخاتمة والآفاق المستقبلية

يهدف هذا الجزء كما تشير تسميته إلى تحديد الغاية من المشروع وتعريف دفتر شروطه الأولي، بالإضافة إلى لمحة عما تم إنجازه فعلياً خلال العمل. يفضل أن يخلو الملخص من المقدمات بحيث لا يتجاوز نصف صفحة. ننصح أيضاً بكتابة نفس الملخص باللغة الإنكليزية (على نفس الصفحة إن أمكن) وذلك بغية التعريف بالعم بشكل أوسع عبر محركات بحث الإنترنت إذا تم وضع التقرير إلكترونياً على الشبكة. (Zhang et al., 2021)

عادةً ما تُطبع صفحة الخلاصة أيضاً على الواجهة الخلفية للتقرير المسماة بالغلاف الرابع (Forth cover)، مما يمكن الآخرين

## المراجع

- [1.] Abagissa, A. T., Saxena, S., & Chandra, J. (2024). *Distilbert-gnn: A Powerful Approach to Social Media Event Detection*. In Review. <https://doi.org/10.21203/rs.3.rs-4193412/v1>
- [2.] Afyouni, I., Aghbari, Z. A., & Razack, R. A. (2022). Multi-feature, multi-modal, and multi-source social event detection: A comprehensive survey. *Information Fusion*, 79, 279–308. <https://doi.org/10.1016/j.inffus.2021.10.013>
- [3.] Aggarwal, C. C., Yu, P. S., Han, J., & Wang, J. (2003). A Framework for Clustering Evolving Data Streams. In *Proceedings 2003 VLDB Conference* (pp. 81–92). Elsevier. <https://doi.org/10.1016/B978-012722442-8/50016-1>
- [4.] Alharbi, A., & Lee, M. (n.d.). *Kawarith: An Arabic Twitter Corpus for Crisis Events*.
- [5.] Deng, S., Zhang, N., Kang, J., Zhang, Y., Zhang, W., & Chen, H. (2020). Meta-Learning with Dynamic-Memory-Based Prototypical Network for Few-Shot Event Detection. *Proceedings of the 13th International Conference on Web Search and Data Mining*, 151–159. <https://doi.org/10.1145/3336191.3371796>
- [6.] Hamoui, B., Mars, M., & Almotairi, K. (n.d.). *FloDusTA: Saudi Tweets Dataset for Flood, Dust Storm, and Traffic Accident Events*.
- [7.] Igartua, M. A., Almenares, F., Redondo, R. P. D., Martín, M. I., Forné, J., Campo, C., Fernández, A., Cruz, L. J. de la, García-Rubio, C., Marínn, A., Mezher, A. M., Díaz, D., Cerezo, H., Rebollo-Monedero, D., Arias, P., & Rico, F. (2020). INRISCO: INcident monitoRing In Smart COMMunities. *IEEE Access*, 8, 72435–72460. <https://doi.org/10.1109/ACCESS.2020.2987483>
- [8.] Li, Q., Chao, Y., Li, D., Lu, Y., & Zhang, C. (2022). Event Detection from Social Media Stream: Methods, Datasets and Opportunities. *2022 IEEE International Conference on Big Data (Big Data)*, 3509–3516. <https://doi.org/10.1109/BigData55660.2022.10020411>
- [9.] McMinn, A. J., Moshfeghi, Y., & Jose, J. M. (2013). Building a large-scale corpus for evaluating event detection on twitter. *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management*, 409–418. <https://doi.org/10.1145/2505515.2505695>
- [10.] Olteanu, A., Castillo, C., Diaz, F., & Vieweg, S. (2014). CrisisLex: A Lexicon for Collecting and Filtering Microblogged Communications in Crises. *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1), 376–385. <https://doi.org/10.1609/icwsm.v8i1.14538>
- [11.] Qazi, U., Imran, M., & Ofli, F. (2020). GeoCoV19: A Dataset of Hundreds of Millions of Multilingual COVID-19 Tweets with Location Information. *SIGSPATIAL Special*, 12(1), 6–15. <https://doi.org/10.1145/3404820.3404823>
- [12.] Saeed, Z., Abbasi, R. A., Maqbool, O., Sadaf, A., Razzak, I., Daud, A., Aljohani, N. R., & Xu, G. (2019). What’s Happening Around the World? A Survey and Framework on Event Detection Techniques on Twitter. *Journal of Grid Computing*, 17(2), 279–312. <https://doi.org/10.1007/s10723-019-09482-2>
- [13.] Schinas, M., Papadopoulos, S., Kompatsiaris, Y., & Mitkas, P. (2018). *Event Detection and Retrieval on Social Media* (Version 1). arXiv. <https://doi.org/10.48550/ARXIV.1807.03675>

- [14.] Suwaileh, R., Elsayed, T., & Imran, M. (2023). IDRISI-D: Arabic and English Datasets and Benchmarks for Location Mention Disambiguation over Disaster Microblogs. *Proceedings of ArabicNLP 2023*, 158–169.  
<https://doi.org/10.18653/v1/2023.arabicnlp-1.14>
- [15.] Wang, X., Wang, Z., Han, X., Jiang, W., Han, R., Liu, Z., Li, J., Li, P., Lin, Y., & Zhou, J. (2020). *MAVEN: A Massive General Domain Event Detection Dataset* (No. arXiv:2004.13590). arXiv. <https://doi.org/10.48550/arXiv.2004.13590>

# الملاحق

ملحق أ

## الدراسة الكميّة

نبين هنا الأرقام والقياسات التي يجب أن يعمل ضمنها النظام