

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

## مشروع تخرج

أعد للنجاح في السنة الخامسة

# الاستفادة من تقنيات التعلم العميق لكشف الأحداث ومراقبتها في وسائل التواصل الاجتماعي

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
8	1.1.1 Typical Workflow for a NED system for unspecified event
9	1.1.1 تعريف المسألة Problem Definition
9	1.1.1 تطور طرق معالجة المسألة
9	1.1.1 المنهجيات والمقاربات لمعالجة المسألة
10	1.1.1 معايير التقييم وقياس الأداء
10	أ. Normalized Mutual Information (NMI)
11	ب. Adjusted Mutual Information (AMI)
11	ت. B-Cubed
11	ب. Precision, Recall
11	1.1.1 مجموعات البيانات المتاحة
12	2.1- مسألة تصنيف الأحداث
12	3.1- مسألة استنباط الموقع الجغرافي من النص
12	1.1.1 تعريف المسألة Problem Definition
12	1.1.1 المنهجيات والمقاربات المتبعة لاستنباط الموقع

12.....	1.1.1 المقاربة المعتمدة على نماذج اللغة الضخمة
12.....	1.1.1 مجموعات البيانات المتاحة
12.....	1.1.1 معايير التقييم وقياس الأداء
13.....	4.1- أنظمة مشابهة
13.....	1.4.1 Live Universal Awareness Map Liveuamap.com
14.....	2.4.1 Emergency and Disaster Information Service (EDIS) rsoe-edis.org
15.....	3.4.1 Global Incidents Map Globalincidentmap.com
17.....	الفصل الثالث
17.....	الدراسة النظرية
18.....	1.1- مقدمة
18.....	3.1- منصات وسائل التواصل الاجتماعي
18.....	2.1- معالجة اللغات الطبيعية
18.....	3.1- ONLINE CLUSTERING
18.....	3.1- GRAPH NEURAL NETWORK
18.....	3.1- HETEROGENEOUS INFORMATION NETWORK
19.....	الفصل الرابع
19.....	الدراسة التحليلية
20.....	1.1- مقدمة
20.....	2.1- مدخلات ومخرجات النظام
20.....	3.1- مخطط حالات الاستخدام
20.....	4.1- السرد النصي لحالات الاستخدام
20.....	4.1- مخطط المكونات
21.....	الفصل الخامس
21.....	المنهجية المقترحة
22.....	1.1- مقدمة
22.....	2.1- المقاربة المتبعة لكشف الأحداث
22.....	3.1- المقاربة المتبعة لتصنيف الأحداث
22.....	4.1- منهجية استنباط الموقع الجغرافي

23	الفصل السادس
23	تصميم النظام
24	1.1- مقدمة
24	2.1- تصميم خدمات النظام
24	3.1- مخطط النظام التصميمي
24	4.1- مخطط النشر DEPLOYMENT DIAGRAM
25	الفصل السابع
25	الأدوات المستخدمة
26	1.1- مقدمة
26	2.1- الهدف من المشروع
27	الفصل الثامن
27	تنجيز النظام
28	1.1- مقدمة
28	2.1- تنجيز الخدمات
28	3.1- مخطط النظام
29	الفصل التاسع
29	تحليل ومناقشة النتائج
30	1.1- مقدمة
30	2.1- معايير التقييم
31	الفصل العاشر
31	اختبارات النظام
32	1.1- مقدمة
32	2.1- اختبار خدمات النظام
32	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، فغدت 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 لحل 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 المقارنة 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. سيركز المشروع على تطبيق تقنيات التعلم العميق التحليل النصوص وجغرافية البيانات المستخلصة من وسائل التواصل الاجتماعي، مع الهدف الرئيسي لاكتشاف الأحداث المحلية مثل الحوادث، الجرائم، والكوارث الطبيعية.

مكونات نطاق المشروع:

1. جمع البيانات:

- سيتم جمع البيانات من وسائل التواصل الاجتماعي، مثل تويتر، تيليغرام، باستخدام واجهات برمجة التطبيقات (APIs) المتاحة لهذه الشبكات أو الزواحف.

2. تحليل النصوص:

- استخدام تقنيات التعلم العميق، لتحليل النصوص وتصنيفها إلى أنواع مختلفة من الأحداث (جرائم، حوادث، ...).

- استخراج الكيانات الأساسية مثل الأماكن والأشخاص المتورطين في هذه 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.

- تطوير آلية لتحويل النصوص الجغرافية إلى إحداثيات جغرافية عبر مرمز 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.

- تقديم 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 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 وتيليجرام والمنصات الإخبارية

4. التخاطب مع APIs ووسائل التواصل الاجتماعي وجمع البيانات منها

5. تمرير البيانات في بيئة دفقية 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 قاد إلى تضخم في البيانات المنشورة من قبل المستخدمين على الانترنت. ومكنت هذه البيانات الضخمة دراسة العديد من المشكلات البحثية، وواحد من مواضيعها المهمة هو كشف الأحداث (Q. 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 "real-world occurrences that unfold time and space"، 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 (Q. 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 (Q. 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 (Q. 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 (Q. 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.

### 1.1.1 Typical Workflow for a NED system for unspecified event

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 وفق البيانات الذي يرد إلى النظام، (2) مرحلة كشف الأحداث، (3) مرحلة مابعد كشف الحدث.

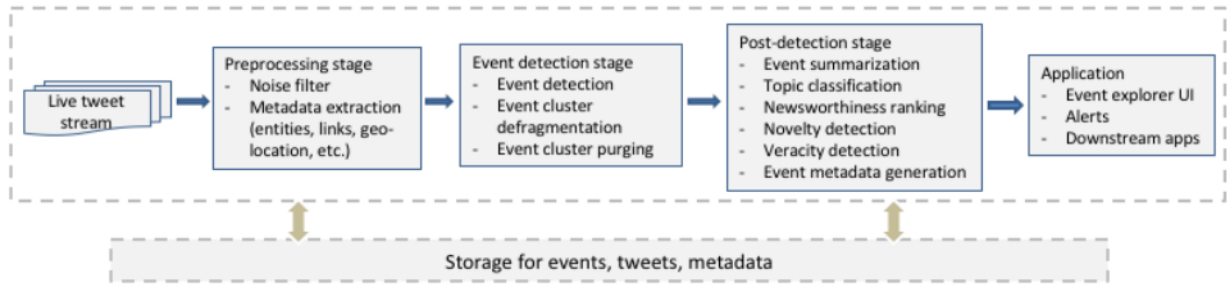


Figure 1 : The typical workflow for a NED system for unspecified events (Q. Li et al., 2022).

ومهمة كل مرحلة هي كما يلي:

♣ مرحلة المعالجة المسبقة Preprocessing Stage، تحوي هذه المرحلة عدة مكونات مثل مكون فلترة الضجيج الذي يقوم بإزالة الرسائل غير المهمة، الإعلانات وماشابه. وتحوي على مكونات لاستخراج الكيانات، المعلومات الجغرافية.

♣ مرحلة كشف الحدث Event Detection Stage، اعتماداً على نوع الحدث المراد كشفه (Specified, Unspecified, RED, NED) قد تقنية الكشف الفعلية (على سبيل المثال، الطرق المعتمدة على العنقدة، 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

1. 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).

2. تعريف: الحدث A Social Event

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

3. تعريف: خوارزمية كشف الأحداث Event Detection Algorithm

من أجل كتلة رسائل  $M_i$ ، تكون 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 تطور طرق معالجة المسألة

#### 1.1.1 المنهجيات والمقاربات لمعالجة المسألة

Methodologies for Event Detection Task					
#	Algorithm	Backbone	Reference	Supervision	Event Type
1	LDA	Topic	7	Unsupervised	Unspecified
2	BiLSTM	DL		Supervised	
3	WMD	WE		Unsupervised	
4	GloVe	WE		Unsupervised	
5	WMD	Similarity		Unsupervised	

6	BERT	PLMs		Unsupervised	
7	SBERT	PLMs		Unsupervised	
8	EventX	CD		Unsupervised	
9	CLKD	GNNs		Supervised	
10	KPGNN	GNNs	(Cao et al., 2021)	Supervised	Unspecified
11	FineEvent	GNNs		Supervised	
12	QSGNN	GNNs	(Ren et al., 2022)	Supervised	
13	ETGNN	GNNs		Supervised	
14	UCLSED	GNNs	(Ren et al., 2023)	Supervised	
15	HCRC	GNNs	(Guo et al., 2024)	Unsupervised	Unspecified
16	RPLMSED	PLMs	(P. Li et al., 2024)	Supervised	
17	HISEvent	CD	(Cao et al., 2023)	Unsupervised	Unspecified
18	ADPSEM	CD	(Yang et al., 2024)	Unsupervised	Unspecified
19	HyperSED	CD	(Yu et al., 2024)	Unsupervised	Unspecified

### 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 تقييمه 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 العقدة (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.

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

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

### ب. Precision, Recall

## 1.1.1 مجموعات البيانات المتاحة

مجموعات البيانات المتاحة لمسألة كشف الأحداث				
#	Dataset	Language	# Events	# Tweets
1	Kawarith	Arabic	7	9000
2	FloDust	Arabic	3	?
3	Event2012	English	503	68,841
4	Event2018	French?	257	64,516
5	MAVEN	English	164	10,242
6	CrisisLexT26	?	6	60,082

7	CrisisMMD	English	7	18,082
8	KBP	?	100	85,569

## 2.1- مسألة تصنيف الأحداث

ن

## 3.1- مسألة استنباط الموقع الجغرافي من النص

### 1.1.1 تعريف المسألة Problem Definition

#### 1.1.1 المنهجيات والمقاربات المتبعة لاستنباط الموقع

#### 1.1.1 المقاربة المعتمدة على نماذج اللغة الضخمة

#### 1.1.1 مجموعات البيانات المتاحة

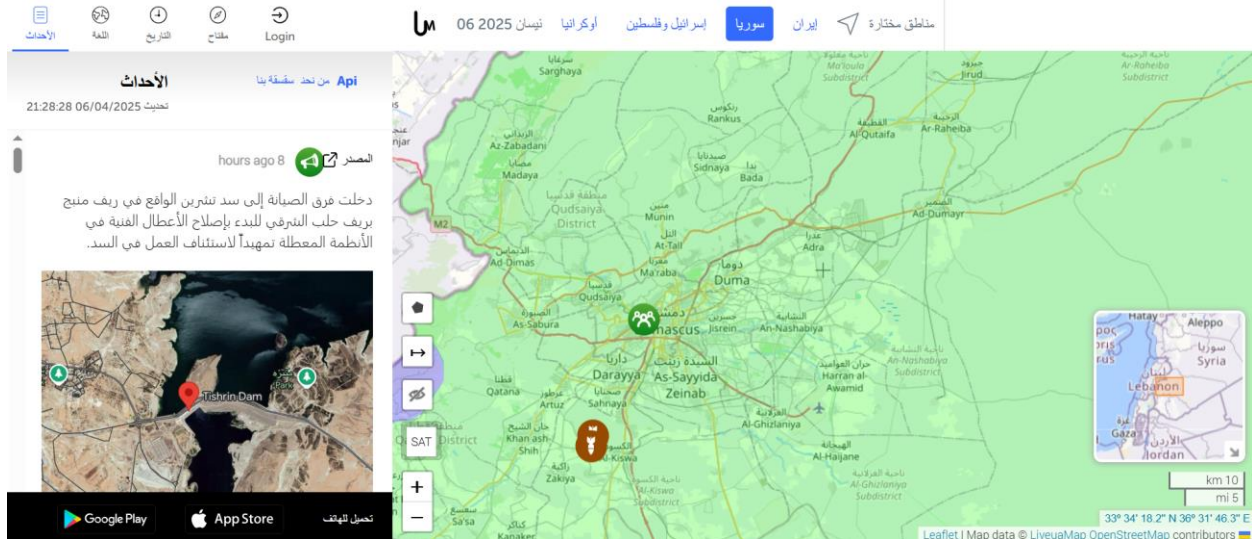
#### 1.1.1 معايير التقييم وقياس الأداء

## 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.

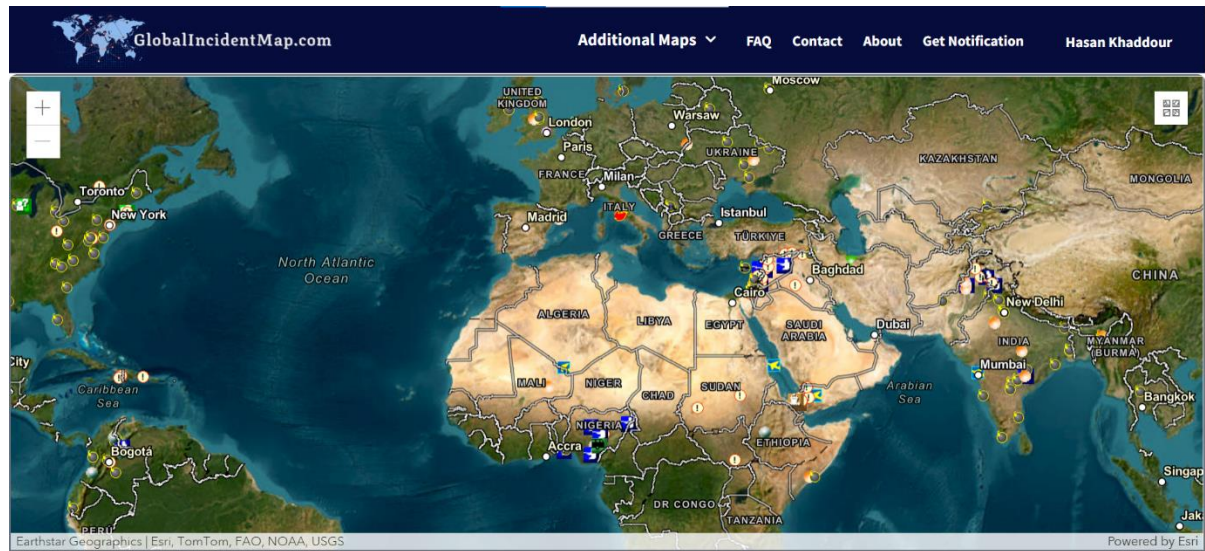
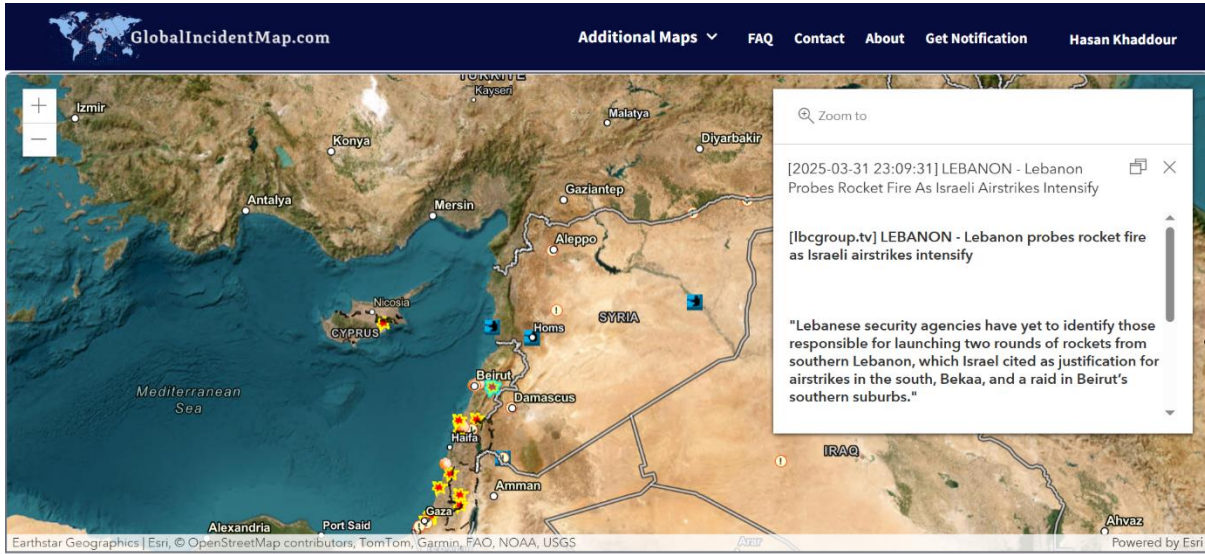


• يوفر تقارير تفصيلية حول أنواع مختلفة من الكوارث مثل:

- 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**

تعتبر

## **3.1- منصات وسائل التواصل الاجتماعي**

## **2.1- معالجة اللغات الطبيعية**

## **3.1- Online Clustering**

## **3.1- Graph Neural Network**

## **3.1- Heterogeneous Information Network**



## الفصل الرابع

# الدراسة التحليلية

يقدم هذا الفصل تحليلاً للمتطلبات التي أوردناها في الفصل الأول.

## 1.1- مقدمة

تعتبر

## 2.1- مدخلات ومخرجات النظام

نطلب

## 3.1- مخطط حالات الاستخدام

يهدف

## 4.1- السرد النصي لحالات الاستخدام

يهدف

## 4.1- مخطط المكونات

ييلب

## الفصل الخامس

# المنهجية المقترحة

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

## 1.1- مقدمة

تعتبر

## 2.1- المقاربة المتبعة لكشف الأحداث

يبل

## 3.1- المقاربة المتبعة لتصنيف الأحداث

يهدف

## 4.1- منهجية استنباط 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.] Cao, Y., Peng, H., Wu, J., Dou, Y., Li, J., & Yu, P. S. (2021). Knowledge-Preserving Incremental Social Event Detection via Heterogeneous GNNs. *Proceedings of the Web Conference 2021*, 3383–3395. <https://doi.org/10.1145/3442381.3449834>
- [3.] Cao, Y., Peng, H., Yu, Z., & Yu, P. S. (2023). *Hierarchical and Incremental Structural Entropy Minimization for Unsupervised Social Event Detection* (No. arXiv:2312.11891). arXiv. <https://doi.org/10.48550/arXiv.2312.11891>
- [4.] Guo, Y., Zang, Z., Gao, H., Xu, X., Wang, R., Liu, L., & Li, J. (2024). Unsupervised Social Event Detection via Hybrid Graph Contrastive Learning and Reinforced Incremental Clustering. *Knowledge-Based Systems*, 284, 111225. <https://doi.org/10.1016/j.knosys.2023.111225>
- [5.] Li, P., Yu, X., Peng, H., Xian, Y., Wang, L., Sun, L., Zhang, J., & Yu, P. S. (2024). *Relational Prompt-based Pre-trained Language Models for Social Event Detection* (No. arXiv:2404.08263). arXiv. <https://doi.org/10.48550/arXiv.2404.08263>
- [6.] 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>
- [7.] Ren, J., Jiang, L., Peng, H., Cao, Y., Wu, J., Yu, P. S., & He, L. (2022). *From Known to Unknown: Quality-aware Self-improving Graph Neural Network for Open Set Social Event Detection* (No. arXiv:2208.06973). arXiv. <https://doi.org/10.48550/arXiv.2208.06973>
- [8.] Ren, J., Peng, H., Jiang, L., Liu, Z., Wu, J., Yu, Z., & Yu, P. S. (2023). *Uncertainty-guided Boundary Learning for Imbalanced Social Event Detection* (No. arXiv:2310.19247). arXiv. <https://doi.org/10.48550/arXiv.2310.19247>
- [9.] Yang, Z., Wei, Y., Li, H., Li, Q., Jiang, L., Sun, L., Yu, X., Hu, C., & Peng, H. (2024). Adaptive Differentially Private Structural Entropy Minimization for Unsupervised Social Event Detection. *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, 2950–2960. <https://doi.org/10.1145/3627673.3679537>
- [10.] Yu, X., Wei, Y., Zhou, S., Yang, Z., Sun, L., Peng, H., Zhu, L., & Yu, P. S. (2024). *Towards Effective, Efficient and Unsupervised Social Event Detection in the Hyperbolic Space* (No. arXiv:2412.10712). arXiv. <https://doi.org/10.48550/arXiv.2412.10712>



# الملاحق

# ملحق أ

## الدراسة الكميّة

نبين هنا الأرقام والقياسات التي يجب أن يعمل ضمنها النظام