مجلة جامعة دمشق للعلوم الهندسية
المجلد41 العدد(1) – 2025: 118-136

Damascus University Journal for engineering sciences
Vo 41 number (1) -2025: 118-136

طريقة جديدة لتسريع بناء الفهرس المكاني R-Tree باستخدام الواصفات النصية للمكان

علاء الدين فريد زين*[1] صلاح الدوه جي[2] محمد إياد الخياط[3]

*[1]. طالب دكتوراه في كلية الهندسة المعلوماتية– جامعة دمشق. adzein@gmail.com

[2]. أستاذ مساعد في كلية الهندسة المعلوماتية– جامعة دمشق. sdowaji@gmail.com

[3]. مدرس في كلية الهندسة المعلوماتية– جامعة دمشق. iyad.khayat@gmail.com

الملخص:

تمثل الفهارس المكانية جزءاً أساسياً في بناء نظم معالجة البيانات المكانية التقليدية، والبيانات المكانية الكبيرة، لأنها –على خلاف طرق التجزئة التقليدية– تستخدم الواصفات ذات الطبيعة الجغرافية كأساس لعملية تجزئة البيانات. إن طريقة الفهرسة المكانية باستخدام بنية المعطيات R-Tree هي أحد الطرق الأساسية المعتمدة في تجزئة البيانات المكانية والتي تعتمد على إضافة معرف الكائن مع مستطيل الحدود الأصغري الخاص به في كل عقدة من أوراق بنية شجرية متعددة المستويات.

يسمح وجود كلمات مفتاحية ذات صيغة نصية واصفة للمكان في مجموعة البيانات المكانية، باستخدام خدمات ويب مثل خدمة المعجم الجغرافي WFS-G، للتغلب على بعض العيوب التي تعاني منها طرق التجزئة باستخدام الفهرس المكاني عند معالجة المضلعات المكونة من أعداد كبيرة من النقاط. يقترح هذا البحث خوارزمية جديدة لتقليص زمن حساب مستطيل الحدود الأصغري، والذي يشكل عاملاً أساسياً في تنفيذ خوارزميات الفهرسة المكانية والتقسيم المكاني، من خلال الاستفادة من الواصفة النصية للمكان في الحصول على مستطيل الحدود الأصغري باستخدام خدمة المعجم الجغرافي. برهنت النتائج العملية قدرة هذا الخوارزمية الجديدة على تقليل الزمن اللازم لبناء الفهرس المكاني R-Tree بشكل واضح عند زيادة عدد النقاط المكونة لمضلعات مجموعة البيانات المكانية عن عتبة قرار مرنة يتم حسابها في مرحلة تهيئة الخوارزمية للعمل.

الكلمات المفتاحية: الفهرس المكاني R-Tree، خدمة المعجم الجغرافي، مستطيل الحدود الأصغري، البيانات المكانية الكبيرة.

تاريخ الايداع:5 /2022/11
تاريخ القبول: 2023/3/13

حقوق النشر: جامعة دمشق – سورية، يحتفظ المؤلفون بحقوق النشر بموجب **CC BY-NC-SA**

Novel method for faster construction of R-Tree spatial index using text-based spatial attributes

Alaa Aldin Fared Zein[*1] Salah Dowaji[2] Mohamad Iyad Al-Khayatt[3]

*[1]. PhD Candidate, Faculty of Information Technology Engineering – Damascus University. . adzein@gmail.com

[2]. Associate Professor, Faculty of Information Technology Engineering – Damascus University. sdowaji@gmail.com

[3]. Assistant Professor, Faculty of Information Technology Engineering – Damascus University. iyad.khayat@gmail.com

Received: 5/11/2022
Accepted: 13/3/2023

Copyright: Damascus University- Syria, The authors retain the copyright under a CC BY- NC-SA

Abstract:

Spatial indexing is a key component of both traditional and big spatial data processing systems because unlike the traditional partitioning methods, they use spatial attributes of objects as a basis for partitioning the data. R-Tree based spatial indexing is one the most used methods of partitioning spatial data. R-Tree is a multi-level tree that uses tuples of identifiers and geographic bounding rectangles as nodes.

Spatial keywords or text-based spatial attributes stored in a spatial data set, allow the use of WFS-G web "Gazetteer" service to overcome some of the shortcomings of spatial indexing, when processing polygons that consists of a very large number of vertices. The new proposed algorithm reduces the time for building minimum bounding rectangles MBRs which is a main factor in creating spatial indexes and spatial partitioning, by using the spatial keywords to obtain minimum bounding rectangles via web "Gazetteer" service. The obtained results prove the capability of the algorithm in reducing R-Tree construction time when the number of vertices of polygons of the data set is above a flexible decision threshold calculated during the initialization phase of the algorithm.

Keywords: R-Tree spatial index, Web Gazetteer service, Minimum Bounding Rectangle, Big Spatial Data

المقدمة:

تتطلب معالجة البيانات المكانية وتنفيذ الاستعلامات عليها بناء فهارس خاصة تأخذ الموقع الجغرافي بعين الاعتبار، تسمى هذه الفهارس بالفهارس المكانية Spatial Index.

بما أن توليد البيانات المكانية يتم بسرعة كبيرة ومن قبل مصادر متعددة وصيغ مختلفة وحجوم كبيرة. وحيث أن العمليات على البيانات المكانية تتطلب خوارزميات ذات تعقيد حسابي عالي. تم التفكير باستخدام الحلول المعلوماتية الخاصة بمعالجة البيانات الكبيرة Big Data في معالجة البيانات المكانية[Li et al.2015]، حيث تعتمد هذه النظم على طرق محسنة لبناء الفهرس المكاني بشكل سريع والاعتماد على هذا الفهرس في تجزئة البيانات المكانية إلى أجزاء تتم معالجتها بشكل تفرعي.

تم تطوير مجموعة من أطر العمل المخصصة للتعامل مع البيانات المكانية الكبيرة من خلال تعديل بنية نظام معالجة البيانات الكبيرة Hadoop و نموذج البرمجة التفرعية Map Reduce ومن الأمثلة على ذلك Hadoop-GIS [Aji et al.2013]، وSpatial Hadoop [Eldawy, Mokbel 2015]. كما تم تطوير نظم مبنية على إطار عمل معالجة البيانات الكبيرة Apache Spark مثل GeoSpark [Yu, Wu, Sarwat 2015]. وكذلك تم تطوير نظم مبنية على نموذج MPI (Message Passing Interface) مثل SpatialMPI [Puri 2019]. وقد تم البرهان على جدوى الحلول السابقة في تحقيق تسريع للأداء عند معالجة البيانات المكانية الكبيرة بالمقارنة مع النظم التقليدية.

تشترك هذه النظم المختلفة في اعتمادها على فهارس مكانية مختلفة–وبشكل أساسي الفهرس المكاني من النوع R-Tree– في بناء آلية الفهرسة والتجزئة للبيانات المكانية. كما تستخدم أيضاً نظم معالجة البيانات المكانية التقليدية مثل Postgre SQL+Post GIS و Oracle Spatial و Python Geopandas ومكتبة PYGEOS الفهرس المكاني R-Tree لفهرسة مجموعات البيانات المكانية.

يتألف الكائن المضلع من مجموعة من النقاط المتصلة لتكوين فراغ هندسي مغلق، حيث تسمى النقاط بالعقد Vertices والخطوط المستقيمة التي تصل فيما بينها بالأضلاع Edges. يقدم هذا البحث طريقة جديدة لتسريع بناء الفهرس المكاني من النوع R-Tree عند التعامل مع مجموعة بيانات مكونة من كائنات من النمط الهندسي المضلع عالي التعقيد (المكون من عدد كبير جداً من النقاط أو العقد Vertices)، من خلال استخدام الواصفات النصية المكانية Spatial Textual Attributes والتي تعرف أيضاً باسم الكلمات المفتاحية المكانية Spatial Keywords، المرافقة للكائنات المكانية، في تسريع عملية حساب مستطيل الحدود الأصغري MBR Minimum Bounding Rectangle والذي يمثل جزءاً جوهرياً من بنية المعطيات R-Tree.

يتضمن هذا البحث تعريفاً بالفهرس المكاني R-Tree، وبخدمة المعجم الجغرافي Gazetteer المستخدمة لتحويل الواصفات النصية المكانية إلى إحداثيات جغرافية، ثم عرض لتصميم الخوارزمية الجديدة بشكل وصفي وبياني وبرمجي، ثم إجراء اختبارات للبرهان على جدوى هذه الخوارزمية الجديدة في تسريع بناء الفهرس المكاني R-Tree.

1. الفهرس المكاني R–Tree:

البنية R-tree [Gutmann 1984]هي بنية معطيات من النوع شجرة Tree، هذه الشجرة من النوع المتوازن الطول Balanced وتحتوي في أوراقها مؤشرات إلى كائنات البيانات، وفي عقدها صفحات من الذاكرة. يسمح استخدام الفهرس R-Tree بتنفيذ الاستعلامات المكانية من خلال زيارة عدد قليل من العقد، وفق بنية ديناميكية تسمح بالإدخال

طريقة جديدة لتسريع بناء الفهرس المكاني R-Tree.....

والحذف والبحث بدون إعادة تنظيم للشجرة. إن نموذج تعقيد عملية البحث ضمن الشجرة هو نموذج لوغاريتمي.

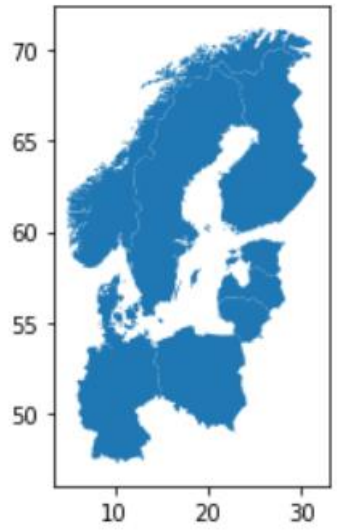
تتألف قاعدة البيانات المكانية من مجموعة من الصفوف المرتبة Tuples التي تمثل الكائن المكاني، ويكون لكل صف معرف فريد، وفي كل عقدة ورقة من الشجرة يكون هناك زوج مكون من فهرس ومعرف فريد، والفهرس هو مصفوفة رباعية القيم تمثل حدود الكائن المكاني (مستطيل الحدود الأصغري MBR Minimum Bounding Rectangle)، وفي كل عقدة من المستوى الأعلى من الشجرة يكون هناك أزواج مكونة من فهرس ومؤشر إلى العقدة الولد بحيث يمثل الفهرس مستطيل الحدود الأصغري للمنطقة التي تغطيها فهارس المستوى الأدنى من الشجرة[Yzelman 2007] .

وبالتالي تعتمد فكرة الفهرسة باستخدام R-tree على تخزين مستطيل الحدود الأصغري لكل كائن في المستوى الأدنى من الشجرة (الأوراق) وثم حساب مستطيل الحدود الأصغري لمجموعة العقد الأبناء لكل عقدة في كل مستوى من الشجرة، وبالتالي يتطلب التنفيذ العملي لهذا الفهرس عملية بناء للشجرة من خلال قراءة كافة الكائنات الموجودة في مجموعة البيانات وحساب مستطيل الحدود الأصغري لكل كائن ثم تحميل جميع الكائنات في الذاكرة لبناء الشجرة Bulk Loading [Zhou, Zhong, Huang 2008].

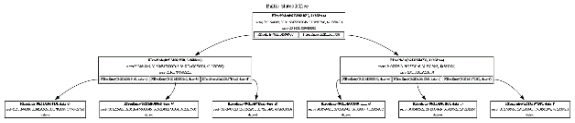
لقد أثبتت الاختبارات أن طريقة التجزئة باستخدام الفهرس المكاني من النمط R-tree حققت أفضل أداء مقارنة بالأنماط الأخرى من الفهارس المكانية لأنها تعتمد في بناء الفهرس على البيانات الموجودة في مجموعة البيانات [Eldawy, Alarabi, Mokbel 2015] . وهذا ما دفع العديد من الباحثين للعمل على تطوير هذا الفهرس المكاني من خلال إضافة مفاهيم جديدة مثل تعلم الآلة إلى الخوارزمية الأصلية [Al-Mamun et al. 2022].

يوضح الشكل (1) مجموعة بيانات مكانية، والشكل (2) يمثل الفهرس المكاني R-Tree الموافق لها ممثلاً بطريقة بيانية،

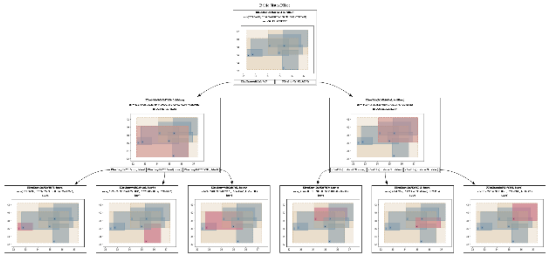
والشكل (3) الفهرس المكاني الموافق لها ممثلاً بطريقة جغرافية.



الشكل (1) مجموعة بيانات مكانية تمثل دول البلطيق



الشكل (2) تمثيل بياني للفهرس المكاني R–Tree لمجموعة بيانات دول البلطيق



الشكل (3) تمثيل جغرافي للفهرس المكاني R–Tree لمجموعة بيانات دول البلطيق

عند دراسة طريقة بناء الفهرس نجد أنها مكونة من ثلاث خطوات أساسية:

- تهيئة بنية المعطيات من خلال بارامترات التهيئة (مثال: عدد الأوراق الأعظمي في العقدة)

طريقة جديدة لتسريع بناء الفهرس المكاني R-Tree.....

- قراءة مجموعة البيانات وحساب مستطيل الحدود الأصغري لكل كائن مع توليد المعرف الفريد لكل كائن
- إضافة العقد (أزواج معرف،مستطيل حدود) إلى الشجرة وتجزئة الشجرة وحساب المستطيل الخاص بالعقد في المستوى الأعلى وصولاً إلى العقدة الجذر.

إن المرحلة المتضمنة حساب مستطيل الحدود الأصغري للكائن المكاني هي خطوة أساسية خاصة عند التعامل مع الكائنات المكانية ذات الشكل الهندسي المضلع حيث يتكون الكائن من عدد من النقاط المتصلة والتي تمثل مدخلات لحساب مستطيل الحدود الأصغري، حيث يتم بناء المستطيل من خلال النموذج الرياضي المبين في الشكل (4):

Calculate MBR for spatial object SO(polygon,{(x1,y1),(x2,y2).....(xn,yn)}):

$\min_X = min(x{:}1..n)$
$\min_Y = min(y{:}1..n)$
$max_X = max(x{:}1..n)$
$max_Y = max(y{:}1..n)$
$MBR = \{\min_X, \min_Y, max_X, max_Y\}$

الشكل (4) نموذج رياضي لحساب مستطيل الحدود الأصغري

يتطلب تنفيذ هذا النموذج خوارزمية ذات درجة تعقيد مرتبطة بعدد النقاط المكونة للكائن المكاني. وبالتالي كلما زاد عدد العقد Vertices المكونة للمضلع كلما زاد زمن حساب مستطيل الحدود الأصغري.

يتم حساب مستطيل الحدود الأصغري في المكتبة GEOS [https://libgeos.org] الواسعة الاستخدام في معالجة البيانات المكانية عبر تنفيذ الشيفرة المبينة في الشكل (5) والتي تتطلب قراءة جميع الإحداثيات المكونة للمضلع قبل بناء المستطيل:


```
void init(double x1, double x2, double y1, double y2)
{
    if(x1 < x2) {
        minx = x1;
        maxx = x2;
    }
    else {
        minx = x2;
        maxx = x1;
    }
    if(y1 < y2) {
        miny = y1;
        maxy = y2;
    }
    else {
        miny = y2;
        maxy = y1;
    }
};

Envelope()
    : minx(DoubleNotANumber)
    , maxx(DoubleNotANumber)
    , miny(DoubleNotANumber)
    , maxy(DoubleNotANumber)
    {};
```

الشكل (5) وظيفة حساب MBR في المكتبة GEOS [https://github.com/libgeos/geos]

وبدراسة الشيفرة المصدرية لبناء شجرة R-tree ضمن نظام Spatial Hadoop [github.org/asel dawy/spatialhadoop2] والمبينة في الشكلين (6) و(7)، نجد أن بناء الفهرس يعتمد على توليد مستطيل الحدود الأصغري MBR من خلال استدعاء مكتبة جافا المعيارية المصممة للتعامل مع المضلعات كأشكال هندسية دون أخذ الواصفة النصية المكانية بعين الاعتبار حيث يتم أيضاً قراءة جميع النقاط المكونة للمضلع من أجل بناء مستطيل الحدود الأصغري.

```
public class Polygon extends java.awt.Polygon implements Shape {

@Override
public Rectangle getMBR() {
  Rectangle2D mbr = super.getBounds2D();
  return new Rectangle((long)mbr.getMinX(), (long)mbr.getMinY(),
      (long)mbr.getWidth(), (long)mbr.getHeight());
}
```

الشكل (6) وظيفة حساب MBR في Spatial Hadoop

```
void calculateBounds(int xpoints[], int ypoints[], int npoints)
    int boundsMinX = Integer.MAX_VALUE;
    int boundsMinY = Integer.MAX_VALUE;
    int boundsMaxX = Integer.MIN_VALUE;
    int boundsMaxY = Integer.MIN_VALUE;

    for (int i = 0; i < npoints; i++) {
        int x = xpoints[i];
        boundsMinX = Math.min(boundsMinX, x);
        boundsMaxX = Math.max(boundsMaxX, x);
        int y = ypoints[i];
        boundsMinY = Math.min(boundsMinY, y);
        boundsMaxY = Math.max(boundsMaxY, y);
    }
    bounds = new Rectangle(boundsMinX, boundsMinY,
                           boundsMaxX - boundsMinX,
                           boundsMaxY - boundsMinY);
}
```

الشكل (7) حساب MBR في مكتبة جافا المعيارية للتعامل مع الأشكال الهندسية

بناء على ما سبق نستنتج أن زمن بناء الفهرس R-Tree يتكون بشكل أساسي من زمن التهيئة وزمن قراءة البيانات وحساب مستطيلات الحدود وزمن إضافة العقد إلى الشجرة.

$$T_{total} = T_{init} + T_{mbr} + T_{insertion}$$

وبالتالي فإن إيجاد طريقة لتقليص زمن بناء المستطيل T_{mbr} له تأثير مباشر على تخفيض زمن بناء الفهرس R-Tree.

2. استخدام الكلمات المفتاحية المكانية في بناء مستطيل الحدود الأصغري:

3.1 المعجم الجغرافي Gazetteer:

المعجم الجغرافي Gazetteer هو قاموس متوفر شبكياً Online يتضمن الأسماء أو المصطلحات المكانية مع التمثيل الهندسي المكاني لها كسمة على الخريطة (شكل هندسي وإحداثيات ونظام إسقاط جغرافي)[NRC 2020]. يمكن استخدام خدمة المعجم الجغرافي للربط بين الأسماء والسمات الجغرافية. أي أنه من الممكن مثلاً تحويل اسم مدينة إلى نقطة أو مضلع مع إحداثيات محددة وفق نظام إسقاط جغرافي معين. كما يستطيع تقديم خدمة البحث والتصفح للبيانات مع وظائف لإضافة بيانات جديدة إلى الخدمة وتجميع عدة مصادر للبيانات الجغرافية ضمن سجلات كاملة لا تحتوي على بيانات مكررة أو غير ذات علاقة. يمكن استخدام الدليل كخدمة مستقلة للإجابة عن استعلامات مكانية كما يمكن استخدامه بشكل ضمني في نظم معلومات أكثر تعقيداً مثل نظم الملاحة ومحركات البحث.

يعمل المعجم الجغرافي على تخزين ما يسمى بالسمات الجغرافية Features والتي يمكن تعريفها بأنها وصف تجريدي لظاهرة في العالم الحقيقي مرتبطة بموقع جغرافي على الكرة الأرضية من خلال إحداثيات ونظام اسقاط جغرافي ومعرف نصي يعبر عن هذا الموقع (اسم مدينة، رمز بريدي). ويمكن إنشاء علاقات بين السمات الجغرافية من قبيل (دمشق موجودة في سورية) أو (لبنان في جوار سورية).

تبرز أهمية المعجم الجغرافي في إعطاء معنى للاسم أو الإحداثية فمثلا يمكن أن تحتوي تغريدة تويتر على احداثيات GPS، ولكن ليتم ربط هذه الإحداثية بموقع ما يجب اللجوء إلى خدمة الدليل الجغرافي لمعرفة اسم الموقع.

إن استخدام وتطوير المعاجم الجغرافية هو جزء أساسي من عمليات الاستعلام والنمذجة المرتبطة بالتمكين الدلالي لخدمات الويب الجغرافية Spatial Semantic web services كما هو مبين في أحدث الدراسات المتعلقة بهذا الموضوع [Kokla, Guilbert, 2020] [Halilali et al. 2022].

3.2 خدمة الويب WFS-G:

خدمة الويب Web Feature Service-Gazetteer هي نسخة معدلة من خدمة الويب Web Feature Service WFS المعيارية. ولتمييزها عن الخدمة الأساسية يتم تسميتها اختصاراً WFS-G [OGC 2012]. تختلف استجابة هذه الخدمة عن WFS بأنها تقدم استجابة معتمدة على المكان المحدد للسمة من خلال الوسم SI _ Location Instance كما تدعم في استعلاماتها العلاقة أب/ابن للسمات الجغرافية أي على سبيل المثال مدينة ضمن منطقة أو ولاية. كما تدعم الخدمة عمليات الفلترة المكانية والغرضية لوضع قيود على نتائج الاستعلام. وقد تم تطوير خدمة WFS-G لتكون خدمة

طريقة جديدة لتسريع بناء الفهرس المكاني R-Tree.....

معيارية قابلة للتحقيق من قبل المبرمجين المختصين في مجال معالجة البيانات. تسمح خدمة المعجم الجغرافي بتنفيذ استعلامات بحث عن اسم أو مصطلح جغرافي معين للوصول إلى التمثيل الجغرافي الهندسي لذلك الاسم كاستجابة من الخدمة. حيث يتم الاستعلام من خلال الاتصال الشبكي بالخدمة كما هو مبين في الشكل (8).

```
<gml:featureMember>
  <SI_LocationInstance
id="CWFID.SI_LOCINST.0.76209.4D91FA52A0661B747E2324020000">
    <guid>{32CE40C6-C3DC-11DE-9C20-00144F233E68}</guid>
    <geographicIdentifier>1809338</geographicIdentifier>
    <alternativeGeographicIdentifiers>
      <alternativeGeographicIdentifier>
        <name>Alcatraz Island Lighthouse</name>
        <nameID>{769EBAEA-CA3E-478F-E044-00144F233E68}</nameID>
        <type>official</type>
        <primary>true</primary>
      </alternativeGeographicIdentifier>
    </alternativeGeographicIdentifiers>
    <position>
      <gml:Point srsName="EPSG:4326">
        <gml:pos srsName="EPSG:4326" srsDimension="3">-122.4219159
37.826039 0</gml:pos>
      </gml:Point>
    </position>
```

الشكل (8) مثال عن استجابة WFS-G

تمتلك هذه الخدمة مجموعة من المناهج مثل GetCapabilities, GetFeature حيث يسمح استدعاء المنهج GetFeature بالحصول على السمة الجغرافية الموافقة للاسم المطلوب في عملية البحث كما هو مبين في الشكل (9)، ومن الممكن أن تكون هذه السمة الجغرافية معبرة عن مستطيل الحدود الأصغري للسمة.

```
<?xml version="1.0" encoding="UTF-8"?>
<GetFeature xmlns="http://www.opengis.net/wfs"
  xmlns:isol9112="http://www.usgs.gov/gnis"
  xmlns:ogc="http://www.opengis.net/ogc"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:gml="http://www.opengis.net/gml" service="WFS" version="1.1.0"
  outputFormat="text/xml; subtype=gml/3.1.1" maxFeatures="100"
handle="">
  <Query typeName="isol9112:SI_LocationInstance" srsName="EPSG:4326">
    <ogc:Filter>
      <ogc:PropertyIsEqualTo>
<ogc:PropertyName>isol9112:SI_LocationInstance/alternativeGeographicId
entifiers/alternativeGeographicIdentifier/name</ogc:PropertyName>
        <ogc:Literal>Hawaii</ogc:Literal>
      </ogc:PropertyIsEqualTo>
    </ogc:Filter>
  </Query>
</GetFeature>
```

الشكل (9) مثال عن طلب WFS Get Feature

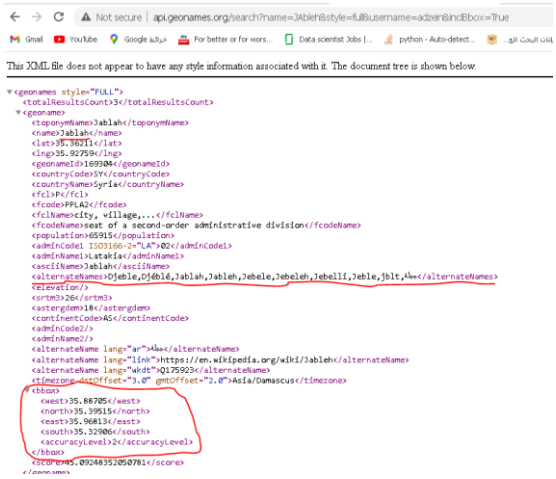
لو أخذنا خدمة Geo Names [www.geonames.org] كمثال على Gazetteer Service لوجدنا أنها تتألف من ملايين السجلات التي تربط بين الأسماء والإحداثيات ضمن بنية شجرية حيث يمكن الوصول إلى البيانات المطلوبة من خلال استدعاء واجهة برمجة التطبيقات API المناسبة للحصول على نتيجة الاستعلام كما هو مبين في الشكل (10).

زين، الدوه جي و الخياط

```
{
  - geonames: [
    - {
        fcodeName: "capital of a political entity",
        toponymName: "Mexico City",
        countrycode: "MX",
        fcl: "P",
        fclName: "city, village,...",
        name: "Mexico City",
        wikipedia: "",
        lng: -99.12766456604,
        fcode: "PPLC",
        geonameId: 3530597,
        lat: 19.428472427036,
        population: 12294193
      },
```

الشكل(10) نموذج عن استجابة خدمة geonames

يمكن إضافة مستطيل الحدود الأصغري إلى استجابة الخدمة Geonames من خلال إدراج هذا الخيار في عنوان URL الطلب مما يؤدي للحصول على استجابة وفق الشكل (11):



الشكل(11) مستطيل الحدود في استجابة خدمة geonames

من التحديات التي تواجه خدمة الدليل الجغرافي غياب مستطيل الحدود الأصغري MBR عن سجلاته أو عدم دقة هذا المستطيل إن وجد خاصة في حال كان يعتمد الدليل على بيانات تطوعية من المستخدمين مما يسبب عدم اتساق في بيانات الخدمات المختلفة. ولمواجهة هذه التحديات تم تطوير خدمة دليل جغرافي تهتم خصوصاً ببناء مستطيل الحدود الأصغري بشكل دقيق لكل سجل من سجلات الدليل الجغرافي مما يسمح ببناء علاقات أب–ابن بين السجلات وكذلك تسهيل عمليات البحث المكاني من خلال التحسين الاحتمالي لعملية بناء المستطيل [Singh, Rafiei 2018].

تم تطوير طريقة لبناء مستطيل الحدود الأصغري من خلال دراسة خدمات الدليل الجغرافي Geo Names ,

OSMNames [Singh 2017]. ومن خلال وجود طريقة لبناء هذا المستطيل وتوسيع الدليل الجغرافي ليشمل مستطيل حدود أصغري أصبح بالإمكان الاستفادة منه في بناء الفهرس المكاني.

تظهر نموذج استجابة مخدم البيانات الجغرافية Geoserver الذي يحتوي على تحقيق مفتوح المصدر للخدمة WFS مع نموذج WFS-G (خدمة ويب المعجم الجغرافي) إمكانية استخدامه في الحصول على مستطيل الحدود الأصغري المرتبط بواصفة مكانية (اسم مكان) محدد كما هو مبين في الشكل (12):


```
{'type': 'FeatureCollection',
 'features': [{'type': 'Feature',
   'id': 'world-administrative-boundaries.128',
   'geometry': {'type': 'MultiPolygon', 'coordinates': ['...']},
   'geometry_name': 'the_geom',
   'properties': {'status': 'Member State',
    'color_code': 'LBN',
    'region': 'Western Asia',
    'iso3': 'LBN',
    'continent': 'Asia',
    'name': 'Lebanon',
    'iso_3166_1_': 'LB',
    'french_shor': 'Liban',
    'bbox': [35.10083, 33.06208, 36.62374, 34.6475]}}],
 'totalFeatures': 1,
 'numberMatched': 1,
 'numberReturned': 1,
 'timeStamp': '2022-10-14T09:55:56.530Z',
 'crs': {'type': 'name', 'properties': {'name': 'urn:ogc:def:crs:EPSG::4326'}},
 'bbox': [33.06208, 35.10083, 34.6475, 36.62374]}
```


الشكل(12) مستطيل الحدود في استجابة WFS

3.3 استخدام الكلمات المفتاحية المكانية في خوارزميات الفهرسة والتجزئة المكانية:

أثبتت الأبحاث أنه من الممكن الاستفادة من الكلمات المفتاحية المكانية الموجودة في مجموعة البيانات في عملية الفهرسة والتجزئة، وتم اقتراح بناء فهرس هجين نصي مكاني للوصول إلى تقسيم عالي الكفاءة وذلك للبيانات المكانية بشكل عام بغض النظر عن حجمها [Kloster 2017]. كما تم اجراء مقارنة للأداء بين الفهارس النصية والفهارس المكانية في تنفيذ الاستعلامات ضمن المستندات التي تحتوي نصوص ذات دلالة جغرافية حيث أثبتت النتائج أن محرك البحث الذي يعتمد على فهرسة نصية ومكانية يحقق أداء أفضل من النظم التقليدية[Christoforaki et al. 2011]، وبناءً على ذلك تم تطوير عدة خوارزميات تستخدم الواصفة المكانية النصية إلى جانب البيانات الجغرافية في بناء فهرس نصي–مكاني ومقارنة أداء هذه الخوارزميات [Chen et al. 2013]. ومن تلك الخوارزميات خوارزمية IR-tree [Li et al. 2011]، والتي تعتمد الواصفة المكانية النصية مع الفهرس المكاني لتحقيق فهرس هجين نصي مكاني والتي تم البرهان على أنها تحقق أداء متفوق في عمليات البحث ضمن المستندات الجغرافية.

3. خوارزمية جديدة لتحسين طريقة بناء الفهرس المكاني R-Tree

4.1 مبدأ عمل الخوارزمية:

تهدف الخوارزمية إلى تحسين طريقة بناء الفهرس المكاني R-Tree من حيث زيادة السرعة مع الحفاظ على المحتوى عند التعامل مع بيانات مكانية من الشكل المضلع، حيث أن هذا النوع من البيانات يزداد تعقيده بزيادة عد النقاط المكونة للمضلع.

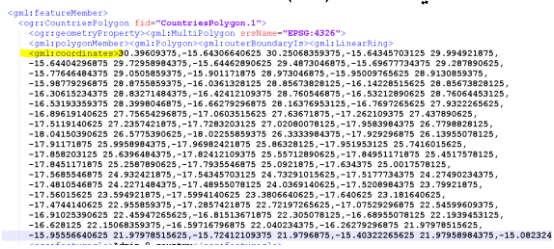
يرتكز مبدأ عمل الخوارزمية على وجود منهجين لبناء مستطيل الحدود الأصغري – وهو العنصر الأساسي الذي يتم تخزينه في عقد الشجرة R-Tree – المنهج الأول يعتمد النموذج الرياضي، والمنهج الثاني يعتمد خدمة المعجم الجغرافي، مع عملية اتخاذ قرار تسمح للخوارزمية باختيار أحد المنهجين عند معالجة كل كائن من مجموعة البيانات المكانية.

في المنهج الأول تتم عملية حساب مستطيل الحدود الأصغري من خلال قراءة جميع الإحداثيات المكونة للكائن المكاني (نقطة، خط، مضلع) وايجاد القيم الحدية القصوى والصغرى للوصول إلى المستطيل المعرف بنقطتين ثنائيتي البعد في حالة مجموعة بيانات ثنائية البعد (نقطة أعلى يسار ونقطة أدنى يمين).

في حال كان الكائن المكاني من النمط المضلع عالي التعقيد (يتألف من عدد كبير نسبياً من العقد) [Zhao et al. 2019]، يصبح حساب المستطيل مسألة مستهلكة للوقت والذاكرة لأنه يتطلب تحميل كل إحداثية من مجموعة البيانات إلى الذاكرة ثم تنفيذ عمليات مقارنة لكل إحداثية وصولاً لإيجاد القيم

طريقة جديدة لتسريع بناء الفهرس المكاني R-Tree.....

الحدية. حيث أننا في البيانات المكانية الكبيرة نتعامل مع مجموعة بيانات كبيرة الحجم مخزنة على الأقراص بصيغة ملف معياري (مثلاً GML-XML أو GeoJSON) وهذه الملفات تحتفظ بقيم جميع الإحداثيات المكونة للكائن المكاني كما هو مبين في الشكل (13).



الشكل (13) تخزين الإحداثيات في ملف GML

في المنهج الثاني يتم بناء خدمة تقوم بتحويل الواصفات المكانية إلى مستطيل حدود أصغري ويمكن أن يتم البناء بعدة طرق:

- **الطريقة الأولى** هي استخدام خدمة ويب متاحة من قبل مزود خدمة من خلال Web Service API مثل الخدمة Geo Names والتي تتضمن ملايين السجلات المتاحة من خلال اشتراك بسيط يتم بموجبه الحصول على هوية ومفتاح Key لاستخدام الخدمة.
- **الطريقة الثانية** هي من خلال استضافة خدمة الويب على مخدم جغرافي محلي مثل المخدم Geo Server وتزويده بقواعد البيانات المناسبة حيث يتم التفاعل مع المخدم من خلال خدمة الويب الجغرافية المعيارية WFS ووفق نموذج البيانات المعياري WFS-G المخصص لتحويل الواصفات النصية المكانية إلى إحداثيات.
- **الطريقة الثالثة** هي من خلال بناء خدمة ويب جديدة وفق النموذج المعياري WFS-G باستخدام لغة البرمجة المناسبة ثم بناء قواعد البيانات التي تحتوي على الأسماء الجغرافية ومستطيلات الحدود الموافقة لها.

4.2 تصميم الخوارزمية

نقترح في الخوارزمية الجديدة الاستفادة من الواصفة النصية المرافقة للكائن المكاني في الحصول على مستطيل الحدود الأصغري من خلال استدعاء خدمة ويب بدلاً من حساب المستطيل لكل كائن من كائنات مجموعة البيانات الكبيرة ومن ثم الاستفادة من مستطيل الحدود في بناء الفهرس المكاني من النوع R-tree.

تهدف الخوارزمية إلى تسريع بناء الفهرس R-Tree لمجموعة بيانات مكونة من مضلعات عالية التعقيد وبالتالي لا بد من تحليل العوامل التي تؤدي إلى زيادة زمن البناء.

إن عملية استدعاء خدمة الويب لها تكلفة زمنية نتيجة عبء الاتصال Communication Overhead مقارنة بالطريقة التقليدية في حساب MBR ذات التكلفة الزمنية الناجمة عن عمليات الحساب المتكررة Computation، وبالتالي يجب أن تكون الخوارزمية الجديدة قادرة على الاختيار بين أحد الطريقتين من خلال استخدام عتبة معينة للقرار. يتم حساب هذه العتبة من خلال حساب الزمن اللازم لبناء مستطيل حدود أصغري لمجموعة ثابتة العدد من النقاط ، وحساب متوسط الزمن اللازم للحصول على استجابة صالحة من خدمة الويب عند ارسال مجموعة ثابتة من الطلبات. كما هو موضح في خطوات المرحلة الأولى من الخوارزمية.

وبالتالي نقترح إضافة عملية اتخاذ قرار إلى خوارزمية التقسيم، حيث تقوم هذه العملية بدراسة عينة من مجموعة بيانات الدخل لمعرفة الشكل الهندسي للكائنات والتي يجب ان تكون من النوع "مضلع Polygon" وكذلك حساب درجة التعقيد لتلك الكائنات المضلعة والتي يمكن تعريفها من خلال معادلة تربط بين درجة التعقيد وعدد العقد المكونة للمضلع.

وأيضاً يجب ان نتأكد من وجود واصفة نصية ذات دلالة مكانية مرافقة لكل كائن للاستفادة منها في بناء الطلب الموجه إلى خدمة الويب.

بناءً على نتيجة عملية اتخاذ القرار إما نتابع العمل بالطريقة التقليدية أو نستخدم الطريقة المعدلة.

لو كان القرار دائماً استخدام الطريقة التقليدية مع جميع كائنات مجموعة البيانات لحصلنا على أسوأ نتيجة متوقعة في هذه الخوارزمية حيث يكون زمن التنفيذ مساوياً الزمن اللازم لبناء الفهرس R-Tree بالطريقة التقليدية مضافاً إليه العبء الناتج عن عملية حساب العتبة واتخاذ القرار.

كما أن الطريقة الجديدة التي تعتمد على تحويل الواصفة المكانية إلى مستطيل حدود أصغري قد تعطي نتائج خاطئة كنتيجة لتشابه الأسماء في بعض الأحيان وهذه يتطلب إضافة خطوة للتحقق Validation من صحة النتيجة إلى الخوارزمية.

تفترض الخوارزمية وجود مجموعة بيانات مكانية من المراد توزيعها على مجموعة من العقد ليتم معالجتها بشكل تفرعي للحصول على النتائج.

دخل الخوارزمية هو مجموعة البيانات وخرج الخوارزمية هو فهرس مكاني من النمط R-tree.

تفترض الخوارزمية وجود طريقتين لبناء مستطيل الحدود الأولى باستخدام الخوارزمية الرياضية التقليدية والثانية باستخدام خدمة تحويل الواصفة المكانية النصية إلى مستطيل حدود.

تعمل الخوارزمية من خلال مراحل متعددة الخطوات:

المرحلة الأولى: مرحلة الدخل والتهيئة وتتضمن:

- الخطوة الأولى هي قراءة ملف الدخل وتحديد نوعه من حيث الشكل الهندسي (نقطي أو مضلع) حيث أن هذه الخوارزمية مخصصة للكائنات المضلعة Polygon، والتأكد من وجود واصفة مكانية مرفقة بالكائن (أي أن هذه الكائنات هي تمثيل لسمة جغرافية وليست مجرد شكل هندسي)
- الخطوة الثانية هي اختيار عينة من الكائنات وحساب الزمن اللازم لبناء مستطيل الحدود الأصغري وتقسيم هذا الزمن على عدد النقاط المكونة للمضلع للحصول على قيمة من الواحدة زمن/نقطة يمكن استخدامها في تقدير زمن بناء المستطيل لأي مضلع مكون من عدد محدد من النقاط n وفق المعادلة: $T_{mb\,r} = t_{point} * n$
- الخطوة الثالثة هي ارسال عدة طلبات تجريبية إلى خدمة الويب التي تقوم بتحويل الأسماء إلى مستطيل حدود وحساب متوسط زمن الاستجابة T_{ws}
- الخطوة الرابعة: إنشاء فهرس مكاني R-Tree فارغ وفق بارمترات افتراضية

المرحلة الثانية: مرحلة التنفيذ وفيها يتم تكرار مجموعة خطوات على جميع الكائنات المكونة لمجموعة البيانات المكانية.

- الخطوة الأولى: قراءة الواصفة المكانية
- الخطوة الثانية: حساب عدد النقاط المكونة للكائن وحساب Tmbr
- الخطوة الثالثة: مقارنة Tmbr مع Tws واتخاذ القرار وفق ما يلي:

Tmbr >Tws: Build MBR using web service method

Tmbr<Tws: Build MBR using traditional method

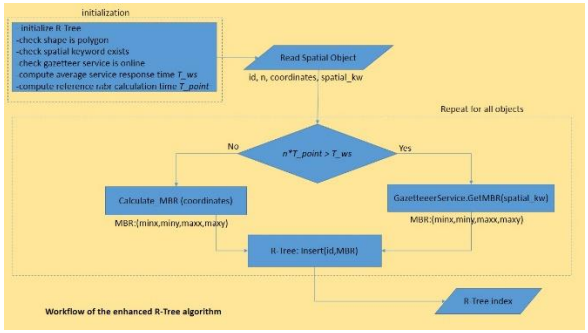
- الخطوة الرابعة بناء المستطيل وفق الطريقة التي تم اختيارها. وفي حال تم بناء المستطيل بطريقة خدمة الويب يتم أخذ نقطة واحدة كعينة من المضلع والتحقق من وقوعها داخل حدود المستطيل للتأكد من صحة المستطيل وإلا نعيد بناء المستطيل باستخدام الطريقة التقليدية.
- الخطوة الخامسة إضافة العقدة الجديدة الموافقة للكائن i إلى الفهرس المكاني R-Tree كزوج من القيم (Idi,MBRi)

المرحلة الثالثة: مرحلة الخرج

وفي هذه المرحلة نكون قد حصلنا على بنية معطيات من النوع R-Tree تتضمن معرفات الكائنات المكونة لمجموعة البيانات مع مستطيل الحدود الخاص بكل منها كأوراق للشجرة، ومستطيلات حدود إجمالية في المستويات الأعلى من الشجرة وصولاً إلى العقدة الجذر.

طريقة جديدة لتسريع بناء الفهرس المكاني R-Tree.....

يمكن في هذه الخطوة تخزين الفهرس بشكل دائم ضمن ملف على القرص من أجل استخدامه عند الحاجة، أو من الممكن الاكتفاء بتخزين الفهرس ضمن بنية المعطيات المناسبة في الذاكرة الرئيسية للحاسب ليتم استخدامه في عمليات الاستعلام والتحليل المكاني اللاحقة. يمكن استخدام المخطط التدفقي المبين في الشكل(14)لتوضيح التصميم:



الشكل(14) مخطط تدفقي لعمل الخوارزمية المقترحة

تمثل شبه الشيفرة البرمجية التالية Pseudo Code خطوات عمل الخوارزمية:

```
//build MBR using traditional method
function calculate_ mbr (x, y)
{
//find_min & find _max are functions of O(n) complexity where n is array length
min x=find _min(x)
max _x=find _max(x)
min _y=find _min(y)
max _y=find _max(y)
mbr=[min _x, min _y,max_ x, max_y]
return mbr
}
//build MBR using web service method
function get_mbr(spatial_kw)
{
//build web api request using keyword
web_api='http://namegeoserver/getFeature?name=sptial_kw'
//get response
response=request(web_api)
//extract mbr (bounding box) from response
mbr=response['bbox']
return mbr
//enhanced method for r-tree building
function build _rtree _enhanced ()
//init phase
ds=read_ spatial _ dataset(file)
//check if polygon
if (ds. geometry .type!='polygon')
//finish algorithm
return -1
//check a sample i for spatial keyword
kw=ds[i].spatial_ keyword _attribute
if (get _ mbr(kw).status!='success')
return -1
//measure average get_ mbr time
repeat 10 times:
t+= get _ mbr (kw).execution _time
time _get _ mbr =t/10
//measure calculate mbr time
x=random(1000 elements)
y=random(1000 elements)
t=calculate _ mbr (x, y).execution _time
time _ calc _ mbr=t/1000
r_ tree =new _rtree (init parameters)
//execution phase
for each so in ds
kw=so .spatial _ keyword_ attribute
n=so. geometry. coordinates. length
//random coordinate for validatation
p=random(so. geometry. coordinates)
threshold=time_ calc _mbr*n
if time _get _mbr <= threshold
mbr=get_ mbr(kw)
//validate
if (p. inside(mbr))
return mbr
//else use traditional method
mbr=calculate _ mbr (so. geometry. x, so. geometry .y)
return mbr
//Insert node to r-tree
R _ tree. insert(so. id, mbr)
//output phase
r_tree. Save _to _disk(file)
```

4.3 الاختبار والنتائج:

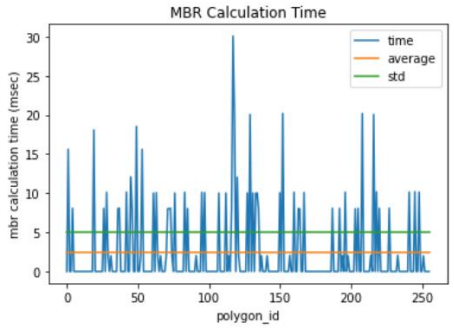
للبرهان على فعالية الخوارزمية المقترحة في تقليل زمن بناء الفهرس R-Tree عند التعامل مع مجموعة بيانات مكانية من النوع المضلع، تم تنفيذ مجموعة من الاختبارات التي تهدف

طريقة جديدة لتسريع بناء الفهرس المكاني R-Tree.....

إلى التحقق التطبيقي من صحة الأسس النظرية التي اعتمدت عليها الخوارزمية.

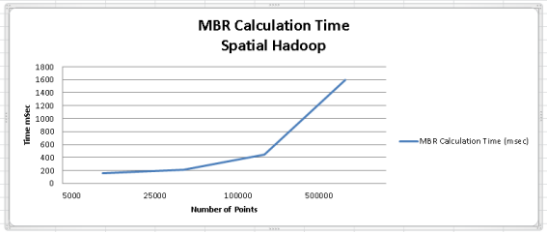
يهدف الاختبار الأول للبرهان على أن زمن عملية بناء مستطيل الحدود الأصغري مرتبط بتعقيد الشكل الهندسي المضلع (عدد النقاط المكونة للمضلع) حيث يزداد الزمن اللازم لبناء المستطيل بزيادة عدد النقاط مما يجعله عاملاً مؤثراً في سرعة بناء الفهرس.

يبين الاختبار الذي قمنا به على مجموعة بيانات مضلعة مختلفة من حيث عدد النقاط المكونة لكل مضلع التباين العالي في سرعة الاستجابة كما هو مبين في الشكل (15):



الشكل(15) زمن بناء مستطيل الحدود لمجموعة بيانات مضلعة

كما تبين التجربة التالية التي قمنا بتنفيذها على نظام SpatialHadoop –وهو أحد نظم معالجة البيانات المكانية الكبيرة– زمن حساب مستطيل الحدود الأصغري لعدة مجموعات بيانات مكانية مختلفة من حيث عدد النقاط. حيث توضح هذه التجربة العلاقة بين زمن بناء مستطيل الحدود الأصغري وعدد النقاط المكونة للكائنات المكانية. يوضح الشكل (16) هذه العلاقة.

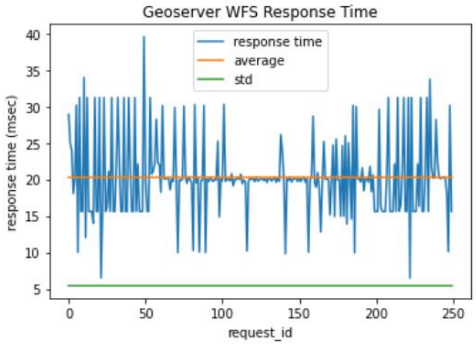


الشكل (16) العلاقة بين عدد النقاط و زمن حساب مستطيل الحدود الأصغري

يبين الشكل (16) زيادة زمن التنفيذ بزيادة عدد النقاط. حيث تأخذ الزيادة منحى غير خطي في البداية نتيجة للعبء المتعلق بتشغيل بيئة العمل، ثم منحى خطي عندما يزداد عدد النقاط عن حد معين، مما يدل على وجود علاقة طردية بين زمن بناء المستطيل وعدد النقاط.

يهدف الاختبار الثاني للبرهان على استقرار زمن استجابة خدمة المعجم الجغرافي وعدم ارتباطه بعدد النقاط المكونة للكائن المكاني حيث تم تنفيذ التجربة بإرسال عدد من الطلبات إلى الخدمة وقياس زمن الحصول على الاستجابة المطلوبة.

يبين الشكل (17) الخط البياني لزمن الاستجابة لخدمة WFS-G Geoserver ومتوسط وتباين زمن الاستجابة ونلاحظ استقرار القيم حول المتوسط، وانخفاض الانحراف المعياري عن المتوسط.



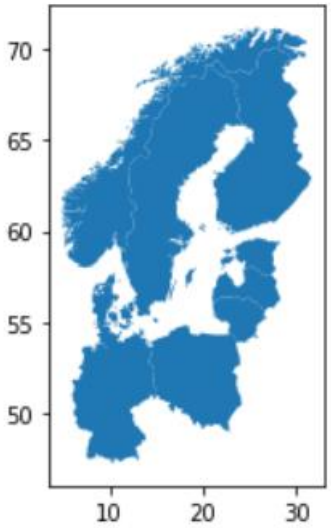
الشكل(17) زمن الحصول على مستطيل الحدود لمجموعة بيانات مضلعة

يهدف الاختبار الثالث إلى البرهان على قدرة الخوارزمية المقترحة على تقليص زمن بناء الفهرس R-Tree عندما تحتوي مجموعة البيانات على مضلعات معقدة مكونة من عدد من النقاط أكبر من عتبة القرار المستخدمة في الخوارزمية.

بغرض تنفيذ الاختبار تم بناء مجموعتين من البيانات المكانية ذات الشكل المضلع (حدود دول البلطيق)، حيث تتميز دول البلطيق بحدود بحرية طويلة ومتغيرة من حيث

طريقة جديدة لتسريع بناء الفهرس المكاني R-Tree.....

وجود الرؤوس والخلجان والجزر ووجود مراكز هامة للنشاط البشري تجعل للبيانات المكانية المرتبطة بها أهمية كبيرة في التطبيقات المختلفة. يمثل الشكل (18) مجموعة بيانات متطابقة جغرافياً مع مجموعة البيانات المبينة في الشكل (1) ولكن في الواقع فإن المجموعة الأولى مكونة من عدد محدود من النقاط المستخدمة في بناء المضلع، أما المجموعة الثانية فهي مكونة من مجموعة أكبر من النقاط لبناء كل مضلع. وبما أن مجموعة البيانات الثانية تمتلك قدرة تمييز مكانية أعلى من مجموعة البيانات الأولى، نطلق على مجموعة البيانات الثانية اسم "دول البلطيق (دقة عالية)".



الشكل (18) مجموعة بيانات دول البلطيق بمضلعات عالية الدقة المكانية

يبين الشكل (19) عينة من جدول الواصفات لمجموعة البيانات ويمكن ملاحظة القيم الكبيرة لعدد النقاط Vertices المكونة لكل مضلع.

	status	color_code	region	iso3	continent	name	iso_3166_1	french_sho	vertices	geometry
6	Member State	NOR	Northern Europe	NOR	Europe	Norway	NO	Norvège	4963027	MULTIPOLYGON (((5.28778 59.21889, 5.28775 59.2...
8	Member State	SWE	Northern Europe	SWE	Europe	Sweden	SE	Suède	1411420	MULTIPOLYGON (((16.43722 56.21139, 16.43721 56...
4	Member State	FIN	Northern Europe	FIN	Europe	Finland	FI	Finlande	1120132	MULTIPOLYGON (((23.70583 59.92722, 23.70577 59...
5	Member State	DEU	Western Europe	DEU	Europe	Germany	DE	Allemagne	968974	MULTIPOLYGON (((14.22555 53.92860, 14.22554 53...
7	Member State	DNK	Northern Europe	DNK	Europe	Denmark	DK	Danemark	907922	MULTIPOLYGON (((11.51389 54.82972, 11.51384 54...
1	Member State	POL	Eastern Europe	POL	Europe	Poland	PL	Pologne	521522	POLYGON ((19.62726 54.46327, 19.62728 54.46326...
2	Member State	EST	Northern Europe	EST	Europe	Estonia	EE	Estonie	424429	MULTIPOLYGON (((22.98485 58.59944, 22.98496 58...
0	Member State	LVA	Northern Europe	LVA	Europe	Latvia	LV	Lettonie	292293	POLYGON ((27.37206 57.53564, 27.37211 57.53564...
3	Member State	LTU	Northern Europe	LTU	Europe	Lithuania	LT	Lituanie	271273	MULTIPOLYGON (((20.98481 55.27655, 20.98477 55...

الشكل (19) عينة من جدول الواصفات لمجموعة بيانات دول البلطيق عالية الدقة المكانية

يبلغ متوسط عدد النقاط المكون لكل مضلع في مجموعة البيانات المبينة في الشكل (1) حوالي 1200 نقطة، والعدد الأعظمي للنقاط في مضلع واحد لا يتجاوز 5000، وفي المقابل يبلغ متوسط عدد العقد في مجموعة البيانات المبينة في الشكل (18) حوالي 3 مليون نقطة والعدد الأعظمي للنقاط في مضلع واحد يزيد عن 6 مليون نقطة.

يهدف الاختبار التالي إلى حساب زمن بناء مستطيل الحدود الأصغري لكل كائن من المجموعتين حيث تم الحصول على النتائج المبينة في الجدول (1)، عند تنفيذ الاختبار على حاسب بمعالج Core i7 الجيل العاشر من 6 نوى و8 GB من الذاكرة الرئيسية مع بناء وظيفة لبناء MBR وفق النموذج الرياضي المبين في الشكل (4) بلغة Python:

الجدول(1) مقارنة زمن بناء MBR لمجموعتي بيانات الاختبار

مجموعة البيانات	زمن بناء MBR
دول البلطيق	0.08 ثانية
دول البلطيق (دقة عالية)	25.7 ثانية

حيث نلاحظ الزيادة الكبيرة في زمن بناء المستطيل.

تلعب مواصفات الحاسب المستخدم دوراً أساسياً في عملية الاختبار، خاصة سرعة المعالج في تنفيذ العمليات الحسابية، حيث أن عتبة القرار المستخدمة في الخوارزمية تعتمد بشكل مباشر على زمن تنفيذ عملية حساب مستطيل الحدود، وتتغير هذه العتبة بتغير مواصفات الحاسب المستخدم. كما أن سرعة وجودة الاتصال بين الحاسب المستخدم في الاختبار، وخدمة الويب التي يتم الوصول إليها عبر شبكة الإنترنت، هي أيضاً عامل مؤثر في حساب عتبة القرار.

نستخدم خدمة الويب Geonames للحصول على مستطيل الحدود الأصغري بالاستفادة من الواصفة النصية المكانية Name الموجودة في مجموعة البيانات فنحصل على النتيجة المبينة في الجدول(2) عند تنفيذ الاختبار باستخدام اتصال انترنت عالي عرض الحزمة:

الجدول(2) زمن الحصول على مستطيل الحدود لمجموعتي البيانات باستخدام Geonames

مجموعة البيانات	زمن بناء MBR
دول البلطيق	7 ثانية
دول البلطيق (دقة عالية)	7.8 ثانية

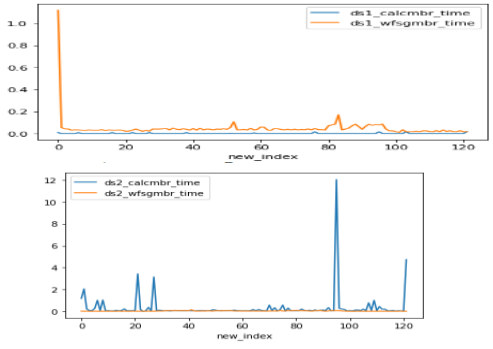
نلاحظ أن زمن بناء المستطيل لم يتأثر بزيادة حجم مجموعة البيانات.

كما نستخدم خدمة الويب WFS المتوفرة في المخدم Geoserver للحصول على مستطيل الحدود الأصغري باستخدام الواصفة النصية المكانية فنجد النتائج المبينة في الجدول (3) عند استخدام مخدم Geoserver على الشبكة المحلية لاستضافة الخدمة WFS:

الجدول(3) زمن الحصول على مستطيل الحدود لمجموعتي البيانات باستخدام Geoserver WFS.

مجموعة البيانات	زمن بناء MBR
دول البلطيق	5.2 ثانية
دول البلطيق (دقة عالية)	5.4 ثانية

يبين الشكل (20) الزمن بالثواني لبناء مستطيل الحدود لكل مضلع من مجموعة البيانات ذات الدقة المنخفضة (ds1) ومجموعة البيانات ذات الدقة العالية (ds2) باستخدام الطريقة الحسابية وباستخدام طريقة خدمة الويب.



الشكل(20) مقارنة زمن بناء مستطيل الحدود بالطريقتين

نلاحظ بالدراسة التحليلية للخطوط البيانية التي تم الحصول عليها أن زمن حساب مستطيل الحدود لكل مضلع وفق الطريقة الحسابية يختلف بين مجموعتي البيانات الأولى والثانية، وهذا واضح من دراسة القيم العظمى للمنحني البياني ds1_calcmbr (من رتبة عشرات الميلي ثانية) ومقارنتها بالقيم العظمى للمنحني البياني ds2_calcmbr (حوالي 12 ثانية). علماً أن ds2 تشتمل على نفس المضلعات الموجودة في ds1 ولكن بصيغة أكثر تعقيداً (عدد أكبر من النقاط المكونة للمضلع). أما بالنسبة للمنحني البياني ds1_wfsgmbr والمنحني البياني ds2_wfsgmbr فنلاحظ أن مجال القيم التي تم الحصول عليها متقارب في المنحنيين (من رتبة عشرات الميلي ثانية) باستثناء القيمة الشاذة التي تم الحصول عليها في بداية المنحني (أكبر من 1 ثانية) والناتجة عن زمن تأسيس الاتصال مع خدمة الويب.

واعتماداً على النتائج التي تم الحصول عليها والتي تبرهن على استقلال زمن بناء مستطيل الحدود عن عدد النقاط المكونة للمضلع عند استخدام خدمة ويب المعجم الجغرافي نقوم ببناء التحقيق البرمجي للخوارزمية المبينة في الشكل (14) باستخدام لغة Python ونقوم بحساب نتيجة وزمن بناء الفهرس المكاني R-Tree باستخدام الطريقة التقليدية وباستخدام الخوارزمية الجديدة، ويتم الاختبار على نفس الحاسب المذكور سابقاً.

يبين الشكل (21) خرج مرحلة التهيئة للخوارزمية والتي يتم فيها يتم حساب عتبة القرار والتي تختلف باختلاف بيئة التنفيذ:

```
Enhanced RTree using Spatial Keywords initilization phase...
1665741955.0669837
[20.96861000000007, 55.67483000000004, 28.235970000000066, 58.08326000000005]
1665741955.1873212
292293
0.12033748626708984
threshold time for one coordinate
4.1170156749251555e-07
sample response time for web service
0.015622854232788086
decision point
choose method two when number of coordinates is more than:
37947.036072608826
```

الشكل(21) مرحلة حساب عتبة القرار

طريقة جديدة لتسريع بناء الفهرس المكاني R-Tree.....

كما يبين الشكل (22) الشيفرة البرمجية لمرحلة اتخاذ القرار للخوارزمية:

```
t1=t.time()
for i in range(len(kw)):
    print(len(poly_x[i]))
    print('object'+str(i)+' out of'+str(len(poly_coords)))
    if (len(poly_x[i]) > int(decision_point)):
        print('using web service method')
        method2_2_s(poly_x[i],poly_y[i],kw[i])
    else:
        print('using traditional method')
        method1_1(poly_x[i],poly_y[i])
t2=t.time()
print('execution time')
print(t2-t1)
```

الشكل(22) – مرحلة اتخاذ القرار

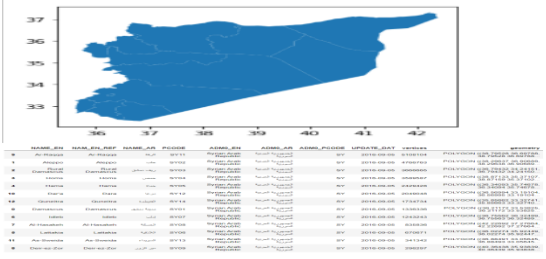
وبتنفيذ التجربة على حاسب الاختبار تم الحصول على النتائج المبينة في الجدول (4):

الجدول(4) زمن بناء R-Tree باستخدام الطريقة التقليدية والطريقة الجديدة المحسنة

الخوارزمية المستخدمة	زمن بناء RTree (دول البلطيق)
RTree مع بناء MBR بطريقة حسابية	26.23 ثانية
RTree مع بناء MBR بالطريقة المحسنة باستخدام خدمة Geoserver WFS	4.19 ثانية
RTree مع بناء MBR بالطريقة المحسنة باستخدام خدمة Geonames	7.86 ثانية

مع الأخذ بعين الاعتبار أن هذا الزمن لا يشمل زمن قراءة مجموعة البيانات والذي يتم بشكل متماثل في الطريقتين التقليدية والمحسنة.

نقوم بتكرار الاختبارات من أجل مجموعة بيانات مبينة في الشكل(23) تمثل الحدود الإدارية لبعض المحافظات السورية بشكل هندسي مضلع عالي الدقة المكانية.



الشكل(23) مجموعة بيانات بعض المحافظات السورية وجدول الواصفات المرافق

زين، الدوه جي و الخياط

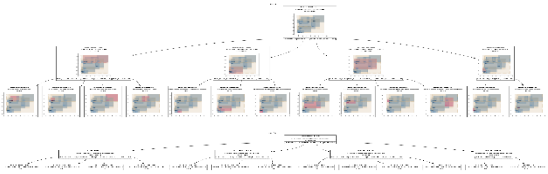
حيث يبلغ متوسط عدد العقد في مجموعة البيانات 2.15 مليون نقطة والعدد الأعظمي للنقاط في مضلع واحد يتجاوز 5 مليون نقطة.

وبتنفيذ التجربة على حاسب الاختبار تم الحصول على النتائج المبينة في الجدول (5):

الجدول(5) مقارنة سرعة بناء R-Tree باستخدام الطريقة التقليدية والطريقة المحسنة باستخدام WFS

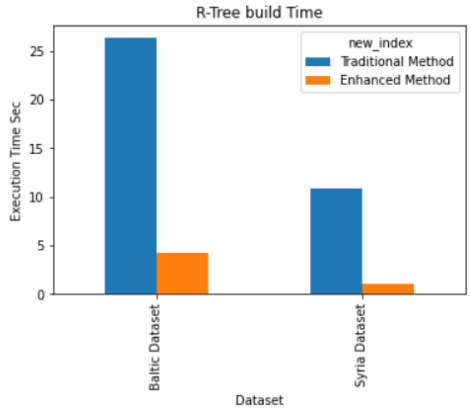
الخوارزمية المستخدمة	زمن بناء RTree (بعض المحافظات السورية)
RTree مع بناء MBR بطريقة حسابية	10.8 ثانية
RTree مع بناء MBR بالطريقة المحسنة باستخدام خدمة Geoserver WFS	1.04 ثانية

نلاحظ أننا حصلنا على نفس الفهرس RTree المبين بيانياً وجغرافياً في الشكل (24) في الحالتين:



الشكل(24) تمثيل بياني وجغرافي للفهرس R-Tree الذي تم الحصول عليه في الحالتين

يمكن تمثيل النتائج التي حصلنا عليها في الاختبارات لمجموعتي البيانات بيانياً بالشكل(25):



الشكل(25) مقارنة سرعة بناء R-Tree باستخدام الطريقة التقليدية والطريقة المحسنة باستخدام WFS

تبين النتائج التحسن الكبير من حيث انخفاض زمن بناء الفهرس R-Tree عند استخدام الطريقة المحسنة.

4.4 مناقشة النتائج:

تبين نتائج الاختبارات التحسين الكبير الذي تم الحصول عليه في زمن بناء الفهرس R-Tree كنتيجة مباشرة لتخفيض زمن بناء مستطيل الحدود الأصغري. تتميز الخوارزمية بأنها تستخدم عتبة قرار مرنة تختلف باختلاف بيئة الاختبار وبالتالي تختار الطريقة المناسبة لحساب المستطيل بناء على معيار كمي منطقي، يجعلها تحقق في أسوأ الأحوال نفس نتائج الطريقة التقليدية. ترتبط درجة التسريع التي تحققها الخوارزمية الجديدة بمواصفات مجموعة البيانات وعدد الكائنات التي تستدعي خدمة الويب من أجل بناء المستطيل، وفي بعض الحالات التي تحتوي فيها مجموعة البيانات على كائن واحد شديد التعقيد يتحقق التسريع كنتيجة لاستخدام طريقة خدمة الويب في حساب المستطيل لهذا الكائن فقط مع استخدام الطريقة التقليدية لباقي الكائنات، لذلك تم تضمين التحقيق البرمجي للخوارزمية عداداً خاصاً لحساب عدد مرات استخدام خدمة الويب وعدد مرات استخدام الطريقة التقليدية.

قد تسبب هذه الخوارزمية بعض الأخطاء عندما يكون هناك التباس في الربط بين الواصفة المكانية النصية والإحداثيات الجغرافية (حالات تشابه الأسماء أو اختلاف التوصيف الجغرافي) فمثلاً مستطيل الحدود الأصغري لدولة الدنمارك يختلف في حالة اعتبار جزيرة غرينلاند جزءاً من الدنمارك عن حالة عدم اعتبارها كذلك. وبالتالي من الممكن في بعض الحالات ان تعطي خدمة الويب نتائج مختلفة للاستعلام عن نفس الاسم مما يتطلب تطوير آليات لتفادي وقوع هذه الأخطاء عند تطوير الخوارزمية مستقبلاً، حيث تتوجه تلك الآليات باتجاه ما يعرف بالفهم الدلالي للمكان Spatial Semantics. كما تتأثر مرحلة حساب عتبة القرار بظروف بيئة التشغيل مما قد يؤدي إلى انحرافات في حساب العتبة مما يتطلب تكرار عملية الحساب عدة مرات واستخدام المتوسط وهذا يزيد من زمن التنفيذ مقارنة بالنتائج الحالية والتي تم فيها اهمال زمن حساب العتبة نظراً لأنه شبه معدوم في التجارب التي قمنا بها مقارنة بالزمن الكلي لتنفيذ الخوارزمية.

4. الخلاصة:

تم في هذا العمل تعريف وتصميم وتنفيذ واختبار خوارزمية جديدة لبناء مستطيل الحدود الأصغري المستخدم في بناء الفهرس المكاني من النوع R-Tree. هذه الخوارزمية موجهة لمعالجة ضعف الأداء الذي ينجم عن حساب مستطيل الحدود الأصغري للمضلعات المكونة من أعداد كبيرة من النقاط. تستفيد هذه الخوارزمية من الواصفة النصية للمكان وخدمة ويب المعجم الجغرافي، وتقوم بالمفاضلة بين اختيار الطريقة الحسابية في بناء مستطيل الحدود الأصغري او استخدام خدمة الويب اعتماداً على عتبة قرار مرنة يتم حسابها بناء على ظروف بيئة التشغيل.

تثبت النتائج التي تم الحصول عليها قدرة هذه الخوارزمية على تحقيق تحسين كبير في سرعة بناء الفهرس R-Tree حيث ترتبط درجة التحسين بعدد النقاط المكونة لمضلعات مجموعة البيانات.

يمكن استخدام هذه الخوارزمية في نظم معالجة البيانات المكانية الكبيرة وتطويرها باتجاه معالجة أي أخطاء متعلقة بالدلالة المكانية للواصفات النصية المرتبطة بالكائنات المكانية.

التمويل: هذا البحث ممول من جامعة دمشق وفق رقم التمويل(501100020595).

References:

[1] S Li, S Dragicevic, F Anton, M Sester, S Winter, J Coltekin, C Pettit, J Haworth, A Stein, T Cheng, Geospatial Big Data Handling Theory and Methods: A Review and Research Challenges, ISPRS Journal of Photogrammetry and Remote Sensing, 115, 119-133. DOI: 10.1016/j.isprsjprs.2015.10.012.

[2] A Aji, F Wang, H Vo, R Lee, Q Liu, X Zhang, JSaltz, Hadoop GIS: A High Performance Spatial Data Warehousing System over MapReduce. Proceedings of the VLDB Endowment, Vol. 6, No. 11 2013 21508097/13/09.

[3] A Eldawy M F. Mokbel, Spatial Hadoop: A MapReduce Framework for Spatial Data, IEEE 2015 IEEE 31st International Conference on Data Engineering Seoul Korea(South) DOI:10.1109/ICDE.2015.7113382.

[4] J Yu, J Wu, M Sarwat, GeoSpark: A Cluster Computing Framework for Processing Spatial Data, 2015 Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems DOI:10.1145/2820783.2820860.

[5] S Puri, Spatial MPI: Message Passing Interface (MPI) for GIS Applications, The Geographic Information Science & Technology Body of Knowledge (2nd Quarter 2019 Edition), John P. Wilson (Ed.). DOI: 10.22224/gistbok/2019.2.6.

[6] A Guttman, RTREEs. A dynamic index structure for spatial searching, University of California Berkeley ACM O-89791-128-8/84/006/0047/84/006/0047
[7] A.N. Yzelman,R-Trees An efficient structure for spatial data management, Master's degree 'Scientific Computing' at the University of Utrecht 2007.

[8] Q Zhou , E Zhong ,Y Huang ,Research on parallel bulk-loading R-Trees based on partition technology of database, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXVII. Part B4. Beijing 2008.

[9] A Eldawy L Alarabi M F. Mokbel, Spatial Partitioning Techniques in SpatialHadoop, .Proceedings of the VLDB Endowment, Vol. 8, No. 12 2015 21508097/15/08.

[10] A Al-Mamun, R Haider, J Wang, W Aref, The "AI+R"-tree: An Instance-optimized R-tree, arXiv:2207.00550 [cs. DB] 2022.

[11] github.com/aseldawy/spatialhadoop2 last accessed 03/2023.

[12] NRC Natural Resources Canada, Gazetteer Service. https://www.nrcan.gc.ca/earth-sciences/geomatics/canadas-spatial-data-infrastructure/standards-policies/8916 last accessed 03/2023.

[13] M S Halilali, E Gouardères, M Gaio, F Devin, Geospatial Web Services Discovery through Semantic Annotation of WPS. ISPRS Int. J. Geo-Inf. 2022, 11, 254. https:// doi.org/10.3390/ijgi11040254 2022
.
[14] M Kokla , E Guilbert, A Review of Geospatial Semantic Information Modeling and Elicitation Approach, ISPRS Int. J. Geo-Inf. 2020, 9, 146; doi:10.3390/ijgi9030146 2020.

[15] Open Geospatial Consortium, Gazetteer Service -Application Profile of the Web Feature Service Best Practice OGC® document: OGC 11-122r1, Publication Date: 2012-02-17.
[16] www.geonames.org, last accessed 03/2023.

[17] S K Singh, D Rafiei, Strategies for Geographical Scoping and Improving a Gazetteer , WWW2018: The 2018Web Conference, April 23–27, 2018, Lyon, France. ACM, New York, NY, USA, Article 4, 10 pages.https://doi.org/10.1145/317 8876.3186078.

[18] S K Singh, Strategies for gazetteer improvement and enrichment, Master's Thesis Department of Computing Science University of Alberta , 2017.

[19] T Kloster, Spatio-textual search on Spark, Master of Science in Informatics IDI Department of Computer Science Norwegian University of Science and Technology 2017.

[20] M Christoforaki , J He, C Dimopoulos, A Markowetz, T Suel ,Text vs. Space: Efficient Geo-Search Query Processing, CIKM'11, October 24–28, 2011, Glasgow, Scotland, UK. Copyright 2011 ACM 978-1-4503-0717-8/11/10.

[21] L Chen, G Cong, C S. Jensen, D Wu ,Spatial Keyword Query Processing: An Experimental Evaluation, Proceedings of the VLDB Endowment, Vol. 6, No. 3 Copyright 2013 VLDB Endowment 2150 8097/13/01.

[22] Z LI, Ken C. K. LEE, B ZHENG, W LEE, D L LEE, X WANG ,IR-Tree: An Efficient Index for Geographic Document Search, IEEE Transactions on Knowledge and Data Engineering. 23, (4), 585-599. Research Collection School Of Information Systems. Institutional Knowledge at Singapore Management University 2011
.
[23] K Zhao, B Jin, H Fan , W Song , S Zhou Y Jiang, High-Performance Overlay Analysis of Massive Geographic Polygons That Considers Shape Complexity in a Cloud Environment,ISPRS Int. J. Geo-Inf. 2019, 8, 290; doi:10.3390/ijgi8070290.