

# KONUMSAL DİZİNLEME YÖNTEMLERİNİN KARŞILAŞTIRILMASI

M. S. Ayhan<sup>1</sup>, H. Sever<sup>2</sup>, H. Gürçay<sup>3</sup>, S. Ak<sup>4</sup>

<sup>1</sup>Ortadoğu Teknik Üniversitesi, Enformatik Enstitüsü, Ankara, msayhan@gmail.com

<sup>2</sup>Çankaya Üniversitesi, Bilgisayar Mühendisliği Bölümü, 06530 Balgat Ankara, sever@cankaya.edu.tr

<sup>3</sup>Hacettepe Üniversitesi, Fen Fakültesi, Matematik Bölümü, 06530 Beytepe Ankara, gurcay@hacettepe.edu.tr

<sup>4</sup>NETCAD, Cyber Plaza, 06800 Bilkent Ankara, serdarak@netcad.com.tr

## ÖZET

Özellikle Coğrafi Bilgi Sistemlerinde ve uygulamalarında konumsal veri türleriyle ilgili sistemler sıkça kullanılmaktadır. Bu karmaşık veri türlerinden oluşan büyük veri kümeleri üzerinde sorguları verimli işleyebilmek için çeşitli konumsal erişim yöntemleri kullanılmaktadır. Bu çalışmada konumsal dizinleme yöntemlerinden olan R-Ağaçlarının, STR-Ağaçlarının, MX-CIF Karesel Ağaçlarının ve GiST R-Ağaçlarının farklı veri modelleri üzerindeki performanslarına odaklanılmıştır.

**Anahtar Sözcükler:** Konumsal Dizinleme, Konumsal Erişim Yöntemleri, Konumsal Veritabanı, R-Ağacı

## ABSTRACT

### COMPARISON OF SPATIAL INDEXING METHODS

Systems concerning of spatial data types have been frequently employed, especially in Geographical Information Systems and applications. Various spatial access methods are used for efficient query processing over these complex and huge data sets. In this study, some of these spatial access methods, R-tree, STR-tree and MX-CIF Quadtree, are dealt and their performances on different data models are focused.

**Keywords:** Spatial Indexing, Spatial Access Methods, Spatial Databases, R-Tree

## 1. GİRİŞ

Bilginin toplanması, depolanması, gerektiğinde işlenmesi, işe yarar hale getirilmesi sistematik işlem adımlarıdır. Bu amaçla ortaya konan sistemlere bilgi sistemleri denir. Bilgi sistemi organizasyonların yönetimsel fonksiyonlarını desteklemek amacı ile bilgiyi toplayan, depolayan, üreten veya dağıtan bir mekanizma olarak tanımlanır. Bilgi sistemi, bilgiyi daha etkin kullanmayı sağlar, doğru karar verme potansiyelini artırır (Yomralıoğlu, 2002).

CBS, coğrafi verileri saklar, geri getirir ve yeni gösterimler için hazırlar, konumsal analizler için araçlar sunar. Kullanıcılara ve karar vericilere karmaşık konumsal problemleri çözme adına etkili araçlar sunarken tatmin edici de bir performans sağlamalıdır (Stefanakis, 1997). Bu alanda ortak bir standart oluşturma çabaları uzun yıllar sonucunda olgunlaşmış ve Open Geospatial Consortium (OGC) ortaya çıkmıştır. Fakat OGC tarafından genel sistem performansına doğrudan ve önemli etkisi bulunan dizinleme yöntemleri ve veri yapıları ile ilgili kesin bir tanımlama veya standart henüz ortaya konulmamıştır. Uygulamalara bağlı olarak geliştirme ekipleri tarafından sistemlere özel gerçekleştirmeler yapılmaktadır. Mevcut sistemlerde uygulanmış olan birçok yöntem arasından da hangisinin hangisine göre ne derecede üstün olduğunu kesin bir şekilde ortaya koymak genellikle yaşanan bu gerçekleştirim farklarından dolayı mümkün olamamaktadır.

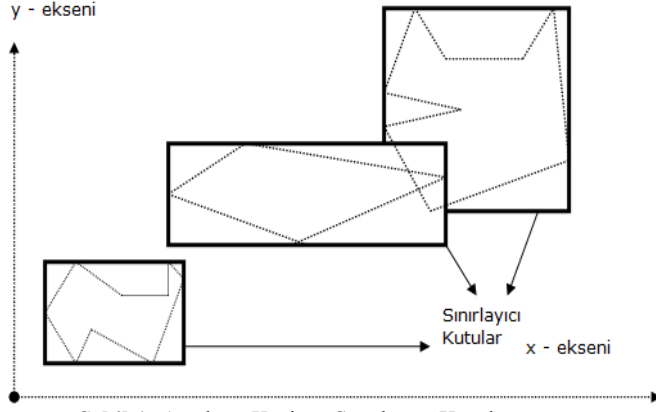
Konumsal dizinleme için çok sayıda yöntem önerilmiştir. Bu sebeple farklı veri konfigürasyonuna ve sorgu türlerine göre konumsal erişim yöntemlerinin performans karşılaştırılmaları yapılmıştır. Bu çalışmada R-ağacı, STR-ağacı, MX-CIF 4'lü ağacı ve GiST R-ağacı adlı dizinleme yöntemlerinin çeşitli konumsal veri dağılımlarına göre karşılaştırılmaları gerçekleştirilmiştir.

## 2. KONUMSAL DİZİNLEME

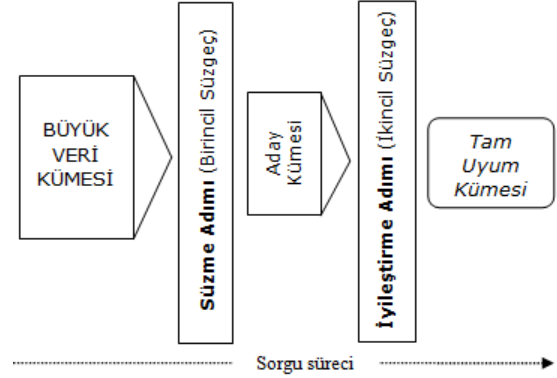
Konumsal verilerin dizinlenmesindeki temel amaç, konumsal nitelikli nesnelerden oluşan büyük kümeler üzerine yapılan sorgularda, sorgu koşuluyla herhangi bir ilişkisi bulunan adayları gruplamaktır. Konumsal dizinleme yöntemleri uzayı ve nesneleri, uzayın bir bölümü veya nesnelerin bir alt kümesi dikkate alınabilecek şekilde kendi içinde organize eder. Bu yöntemlerin temelinde yatan ana fikir konumsal nesnelere yaklaşımlarda bulunmaktır. Bu sayede geometrik özellikleri bulunan konumsal nesneler geometrik karmaşıklıklarından biraz olsun soyutlanarak daha basit şekillerde görülebilmektedir. Bu yaklaşımların en genel ve geçerli örneği *sınırlayıcı kutu* mantığıdır (Şekil 1). Sınırlayıcı kutu, konumsal bir nesneyi sınırlayan, kenarları koordinat eksenlerine paralel olarak çizilebilecek en küçük diktörtgen olarak tanımlanır.

Konumsal sorgu, çözümlendikten sonra değerlendirilmesi için izlenen genel yöntem iki aşamalıdır (Şekil 2).

- i. **Süzme Adımı:** Nesneleri daha basit şekillerde ele alabilmek için gerçek geometrileri yerine yaklaşımlarından faydalanarak oluşturulan *dizin yapıları* sayesinde ilgisiz nesnelere elenmekte ve konumsal sorgunun sonucu açısından potansiyeli yüksek olan *aday nesneler* belirlenmektedir.
- ii. **İyileştirme Adımı:** Bu adımda süzme adımıyla aday olarak belirlenen nesnelerin gerçek geometrileri üzerinden detaylı karşılaştırmalar yapılır. Adaylar kümesi içerisinde, gerçekten aranan nesneler, karmaşık geometrik hesaplamalar sonucunda bulunduktan sonra tam uyum sağlayan nesnelerin kümesi elde edilmiş olur. Bir başka ifadeyle önceki adımda bulunan sonuç iyileştirilmiş olmaktadır.



Şekil 1: Ayrık ve Kesişen Sınırlayıcı Kutular



Şekil 2: İki aşamalı sorgu modeli

### 3. İNCELENEN YÖNTEMLER

Konumsal sorgulama için kullanılan veri yapıları genelde B-ağaçlarıdır. Bu bize verilere I/O işlemleri için logaritmik olarak erişmemizi sağlar. Ayrıca B-ağaçları bize verilerin bir total sıralamasını vererek verilerin birbirleriyle karşılaştırılabilirliğini sağlar. Aşağıda bu çalışmada kullandığımız bazı B-ağaçlarına kısaca değinilmiştir.

#### 3.1 R-ağacı

Veri güdümlü konumsal erişim yöntemlerinin başında gelen R-ağaçları Antonin Guttman tarafından öne sürülmüş, B-ağaçlarını temel alan dengeli ağaçlardır (Guttman, 1984). Yapıları dikdörtgenlerin düzlemdeki dağılımına göre şekillenir, hiyerarşiktir. R-ağaçları dikdörtgenlerin içirme ilişkilerine dayanarak bir düzenlenir. Bu düzen içerisinde her bir düğüm, bütün çocuklarını içeren en küçük sınırlayıcı kutu ile ilişkilendirilir. Hedeflenen bir dikdörtgene erişmek için R-ağacı kökten yapraklara doğru izlenen bir yolla taranır ve her düğümde, o düğümün sınırlayıcı kutusuna bakılarak aranan kayıtla örtüşme durumu sınanır. R-ağacının performansı dizinlenen nesne sayısına göre logaritmiktir. R-ağacıyla kurulan bir dizin yapısı tamamen dinamikdir. Ekleme ve silme işlemleri sürekli olarak yinelenabilir, sorgular arasında gerçekleşebilir ve herhangi periyodik bir düzenleme gerektirmez (Guttman, 1984).

Sağladığı dinamik yapının yanında ne yazık ki geride kaldığı noktalar da vardır. Alan etkinliği açısından düşünüldüğünde yoğun ekleme ve silme işlemleri sonrasında düğümlerde tam doluluk oranı sağlanamadığında performans düşüklükleri oluşmaktadır. Tam doluluk oranı yakalamak adına düğümler yeniden düzenlenmeli ve yoğunlaştırılmalıdır.

Dinamik bir ortamda R-ağacı %50 civarında bir doluluk oranı belirler, deneysel çalışmalar ise ortalama olarak bu oranın %70 olduğunu göstermektedir (Kamel, 1993). Durağan bir ortamda bu oran daha da iyileştirilebilir. Doluluk oranının artması sayesinde ağaçtaki boşluklar azalacak ve bu boşluklardan dolayı derinliği artan ağaç küçülecek, sıklaşacaktır. Bu sayede de arama performansı tepki süresi açısından iyileşecektir. Ağacın yoğunlaştırılması için veriler eklenmeden önce bir takım *ön işlemlere* tabi tutulur.

#### 3.2 STR (Sort-Tile-Recursive) Yoğunlaştırma Algoritması

STR algoritması basit ve etkili bir çözümle kendisinden önceki yoğunlaştırma algoritmalarına oranla disk erişimlerinde %50'ye varan üstünlük sergilemektedir (Leutenegger, 1997). N tane dikdörtgen, kapasitesi M olan  $\lceil N/M \rceil$  tane yaprağa yerleştirilecektir. Yaprak sayısına P dersek, yatayda ve dikeyde sayıları eşit ve  $\sqrt{P}$  olan dilimlere ayrılan düzlemde, N tane dikdörtgen sıralanıp bu dilimlere yerleştirilecektir. İlk olarak dikdörtgenler merkezi noktalarının x koordinatına göre sıralanır ve S tane dikey dilim oluşturulur. Dikey dilimleri belirledikten sonra y koordinatına göre sıralama yapılır ve M tane dikdörtgen barındıran gruplara ayrılır. Bu aşamada

yoğunlaştırılmış (sıkıştırılmış) R-ağacının yaprakları açığa çıkar. Dikdörtgenler ön işlemde geçirilip yaprakların içerikleri düzenlendikten sonra öz yinelemeli olarak yukarı seviyeler oluşturulur.

### 3.3 MX-CIF Dörtlü Ağacı

Dörtlü ağaç kavramı, veri uzayını öz yinelemeli olarak dörderli parçalara bölümleyerek, hiyerarşik bir yapı kurma ilkesine dayanan veri yapılarına verilen genel isimdir. MX-CIF 4'lü ağacı, Gershon Kedem tarafından öne sürülen 4'lü ağaç yapısıdır. MX-CIF 4'lü ağacında her bir R dikdörtgeni, kendisini bütünüyle örten, kapsayan en küçük hücreyle ilişkilendirilir. Dikdörtgenler yaprak veya yaprak olmayan ara düğümlerle ilişkilendirilebilir. Özyinelemeli olarak gerçekleşen bölümlenme süreci, bir düğüm hiç bir dikdörtgen kapsamayınca kadar devam eder. Veya bir diğer seçenek olarak, bu sonlanma koşulu bölümlenen hücre ebatlarının belirli bir eşik düzeyinin altına inmesi durumuna bağlanabilir. Bu eşik değeri genellikle beklenen bir dikdörtgen ebadına göre seçilir (Kedem, 1982).

## 4. YAPILAN ÇALIŞMA

### 4.1 Veri Üretimi

Farklı veri dağılımlarının dizinleme yöntemlerinin performansına etkisini değerlendirebilmek adına sentetik veri üretimi yapılmaktadır. Üretilen veriler ilk etapta sınırlayıcı kutulardır ve temel olarak iki farklı yaklaşım benimsenmiştir. İlk olarak sabit bir istatistiksel dağılıma bağlı kalınarak veri uzayına dağıtılan sınırlayıcı kutular bulunmaktadır. Sabit ve standart dağılımlar sayesinde dağılımlara göre veri artışının dizinleme yöntemlerine etkisinin değerlendirilmesi amaçlanmaktadır. Veri üretiminde kullanılan temel iki dağılım fonksiyonu *düzgün dağılım* ve *Gauss dağılımı*dır. Bu dağılımlara göre kutu üretimi yapılırken kullanılan parametreler anlamlarıyla ve kullanılan değerleriyle birlikte Tablo 1'de ve Tablo 2'de verilmiştir.

| Parametre | Anlamı                                                 |
|-----------|--------------------------------------------------------|
| d         | Kutu ebatlarının çarpanı (büyütme, küçültme değişkeni) |
| N         | Üretilen sınırlayıcı kutu sayısı                       |
| Min_x     | Veri düzlemindeki en küçük x koordinatı değeri         |
| Min_y     | Veri düzlemindeki en küçük y koordinatı değeri         |
| Max_x     | Veri düzlemindeki en büyük x koordinatı değeri         |
| Max_y     | Veri düzlemindeki en büyük y koordinatı değeri         |
| VDF       | Veri dağılım fonksiyonu                                |

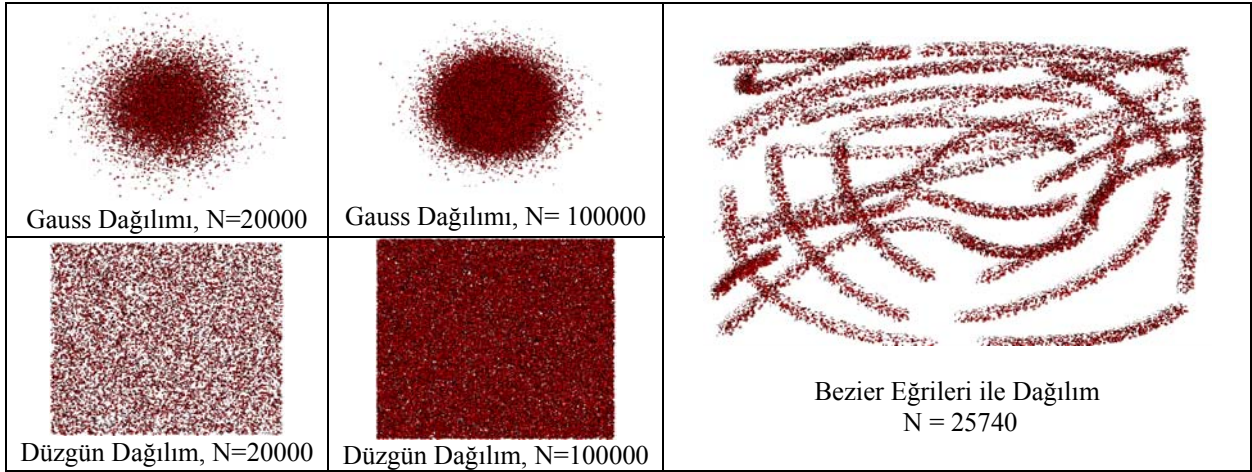
**Tablo 1:** Sınırlayıcı Kutu Üretim Parametreleri

| VDF                 | Gauss Dağılımı |              | Düzgün Dağılım |              |
|---------------------|----------------|--------------|----------------|--------------|
| N                   | 20000          | 100000       | 20000          | 100000       |
| Min_x<br>,<br>Min_y | 0 , 0          | 0 , 0        | 0 , 0          | 0 , 0        |
| Max_x<br>,<br>Max_y | 500 ,<br>500   | 500 ,<br>500 | 500 ,<br>500   | 500 ,<br>500 |
| d                   | 0.01           | 0.01         | 0.01           | 0.01         |

**Tablo 2:** Kutu üretiminde kullanılan değerler

Standart veri dağılımlarının yanında rastsallığın da konumsal erişim yöntemlerine etkisini görebilmek ve istenen tarzda bir şehri veya coğrafi veri dağılımına sahip herhangi bir bölgeyi modelleyebilmek için Bezier eğrilerinden faydalanılmıştır. Bu sayede Bezier eğrileri ile veri uzayında istenen bölgelere ana hatlar çizilmektedir. İkinci aşama olarak ise bu hatlar üzerinde düzgün dağılıma sahip dikdörtgenler oluşturulmaktadır.

Sınırlayıcı kutular üretildikten sonra, kutular içerisine her bir köşesi, sınırlayıcı kutunun farklı kenarına gelecek şekilde rastgele dörtgenler yerleştirilmektedir. Bu sayede istenen dağılıma sahip rastgele geometrik nesneler elde edilmektedir (Şekil 3). Veriler 2 boyutlu uzayda bir düzlem üzerinde dağılmaktadır. İstatistiksel dağılımlara göre üretilen veriler 500 x 500 ebatlarında bir düzlemde bulunmaktadır. Bezier eğrileri yardımıyla üretilen veriler ise 1024 x 768 ebatlarında bir düzlemde.



Şekil 3: Veri kümelerinin kuşbakışı görünümü

#### 4.2 PostgreSQL VTYS ve PostGIS

Temelleri 1977’de Berkeley üniversitesinde atılan PostgreSQL, ilişkisel modellere dayanan ve standart SQL sorgu dilini destekleyen bir VTYS’dir. Güncel sürümlerinde yeni nesil özellikler sunan PostgreSQL, nesne yönelimli tasarım ilkelerini de kapsayan yapısıyla genişleyebilme yeteneğine sahiptir. Bununla birlikte basit veri tiplerine ek olarak sonradan tanımlanabilen veri tiplerini kullanabilmeyi de sağlar. Bu sayede yeni nesil VTYS örneklerinin çoğunda görülen nesnesel ve ilişkisel olma özelliklerini harmanlar.

PostgreSQL VTYS ile beraber kullanılmak üzere geliştirilen PostGIS, konumsal veritabanı teknolojileri ile ilgili açık kaynak kodlu bir projedir ve mevcut VTYS’ye coğrafi nesneler için konumsal destek vermektedir. Bu destek sayesinde, PostgreSQL, coğrafi bilgi sistemleri için güçlü bir VTYS adayı olmaktadır. PostGIS OGC tanımlarına ve kısıtlamalarına uygundur (RAPOR1). OGC standartlarına göre konumsal nesnelerin gösterimi için 2 yol vardır:

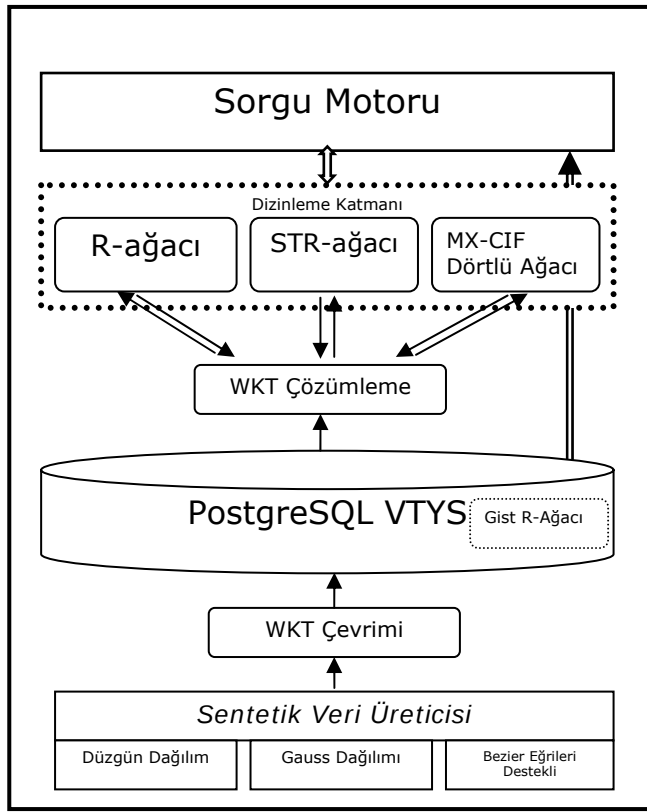
- ◆ WKT (Well-Known Text) : *Düzgün biçimli metin* olarak çevrilebilecek bu kavram konumsal nesnelerin metinsel bir biçimde gösterimidir.
- ◆ WKB (Well-Known Binary) : *Düzgün biçimli ikili kod* olarak çevrilebilecek bu kavram konumsal nesnelerin metinsel gösterimlerinin ikili halde kodlanmış karşılığıdır.

#### 4.3 Deney Düzenegi

Bu çalışma için gerekli olan testler genel teknik özellikleri aşağıda belirtilen bilgisayarda yapılmıştır.

|            |                                           |                 |                              |
|------------|-------------------------------------------|-----------------|------------------------------|
| İşlemci    | : Intel Pentium M 1.60 Ghz, 2 MB önbellek | Ana Hafıza      | : 512 MB DDR PC2700, 333 MHz |
| Sabit Disk | : 40 GB 4200 RPM, 8 MB önbellek           | İşletim Sistemi | : Windows XP Service Pack 2  |

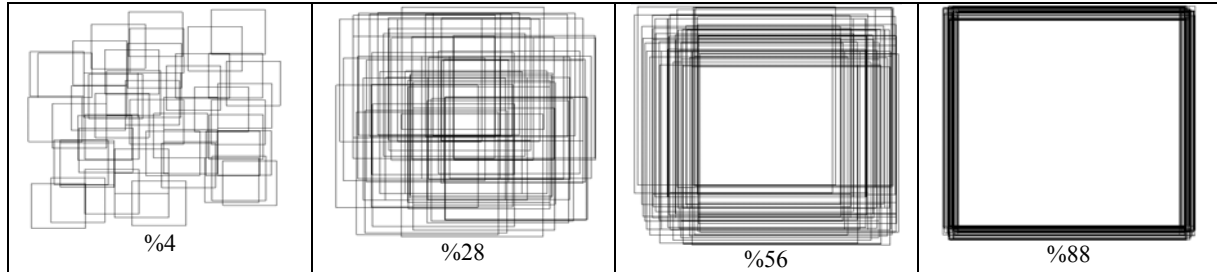
Dizinleme katmanı olarak VTYS üzerine yerleşen yöntemler ile PostGIS eklentisiyle PostgreSQL VTYS’nin kendi bünyesinde yer alan Gist R-ağacının tepkilerini karşılaştırabilmek için, VTYS ile bütünleşik olan izin yapısı da aynı testlere tabi tutulacaktır. İlk olarak düzgün dağılım, Gauss dağılımı ve Bezier eğrileri yardımıyla veri üretimi gerçekleştirildikten sonra bu geometrik veriler düzgün biçimli metinler (WKT) halinde veritabanına girilir. Fakat içsel olarak veritabanında ikili kodlar halinde tutulur. Veritabanında tutulan her nesnenin ayırt edici bir numarası vardır. Dizin yapıları kurulurken geometrik nesnelerin sınırlayıcı kutularına ihtiyaç duyulduğundan, dizinleme katmanı ile VTYS arasında düzgün biçimli metinleri çözümleyen bir mekanizma bulunmaktadır. Bu çözümleyici sayesinde geometrik nesnelerin sınırlayıcı kutuları hesaplanıp izin yapıları kurulmaktadır. Dizin yapısı içerisinde verilerin nesne anahtarları ve sınırlayıcı kutuları tutulmaktadır.



Şekil 4: Deney düzeneği genel yapısı

| Pencere Büyüklük Numarası (PBN) | Pencere Büyüklük Yüzdesi | Pencere Sayısı |
|---------------------------------|--------------------------|----------------|
| 1                               | %4                       | 50             |
| 2                               | %8                       | 50             |
| 3                               | %12                      | 50             |
| 4                               | %16                      | 50             |
| 5                               | %20                      | 50             |
| 6                               | %24                      | 50             |
| 7                               | %28                      | 50             |
| 8                               | %32                      | 50             |
| 9                               | %36                      | 50             |
| 10                              | %40                      | 50             |
| 11                              | %44                      | 50             |
| 12                              | %48                      | 50             |
| 13                              | %52                      | 50             |
| 14                              | %56                      | 50             |
| 15                              | %60                      | 50             |
| 16                              | %64                      | 50             |
| 17                              | %68                      | 50             |
| 18                              | %72                      | 50             |
| 19                              | %76                      | 50             |
| 20                              | %80                      | 50             |
| 21                              | %84                      | 50             |
| 22                              | %88                      | 50             |
| 23                              | %92                      | 50             |
| 24                              | %96                      | 50             |
| 25                              | %100                     | 50             |

Tablo 3: Sorgu penceresi büyüklüğü tablosu



Şekil 5: Örnek pencere yerleşimleri

#### 4.4 Test Senaryoları

Dizinleme yöntemlerini birbirleriyle karşılaştırabilmek için ortak bir test senaryosu gerekmektedir. Üretilen sentetik veriler, dizin yapıları ve Şekil 4'teki yapı çerçevesinde veriler üzerinde pencere sorguları yapılacaktır. Karşılaştırmalarda esas olarak sorgu işleme sürecinde birincil süzgeçten (süzme adımı) dönen aday küme boyutları ve bunun için gereken zamanlar kıyaslanacaktır.

Belirtilen veri kümeleri üzerinde yapılacak sorgularda kullanılmak üzere 25 farklı boyuttan 50'şer adet farklı pencere düzgün dağılıma sahip olacak şekilde üretilmektedir. Bu pencereler uygulanan senaryolarda dinamik olarak yaratılır ve rastgele yerleştirilirler. Tablo 3'te pencerelerin büyüme adımları ve oranları, Şekil 5'te ise örnek pencere yerleşimleri verilmiştir.

## 5. SONUÇLAR

Test senaryoları uygulanmadan hemen önce veritabanındaki veriler doğrultusunda dinamik olarak yüklenen dizinler VTYS üzerinde dizinleme katmanı oluşturmaktadır. Sonuç grafiklerinde sorgu başarımı kıyaslaması yapılırken esas alınan temel iki nokta aday kümesi oluşum zamanı (*milisaniye*) ve aday kümelerinin eleman sayılarıdır. Bütün testlerde R-ağacı ve STR-ağacının dallanma faktörü 10 (on) olarak sabit tutulmuştur, her bir düğümün 10 adet

çocuğu olabilmektedir. Sonuçlara ilişkin verilen grafikler ve tablolarda, veri uzayına göre 25 farklı oranda üretilen 50 pencerenin ortalaması alınmıştır. Sorgu zamanlarının ölçümleri *milisaniye (ms)* olarak verilmiştir.

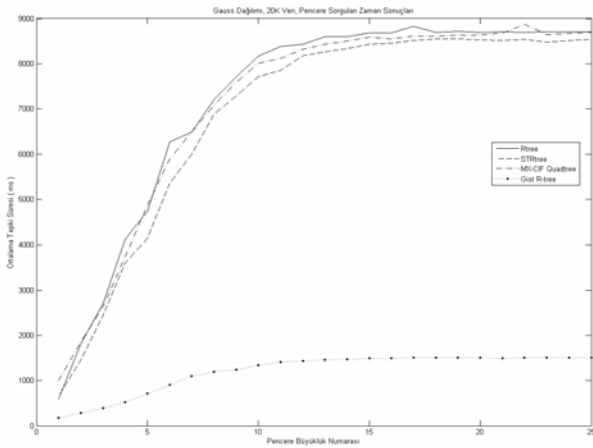
### 5.1 Gauss Dağılımı Sonuçları

|                           | Gauss Dağılımı, 20K veri |           |                  | Gauss Dağılımı, 100K veri |           |                  |
|---------------------------|--------------------------|-----------|------------------|---------------------------|-----------|------------------|
|                           | R-Ağacı                  | STR-Ağacı | MXCIF 4'lü Ağacı | R-Ağacı                   | STR-Ağacı | MXCIF 4'lü Ağacı |
| <b>Derinlik</b>           | 5                        | 5         | 18               | 6                         | 6         | 23               |
| <b>Oluşum Zamanı (ms)</b> | 6298,13                  | 6458,75   | 7887,5           | 62991,25                  | 32743,75  | 39384,38         |

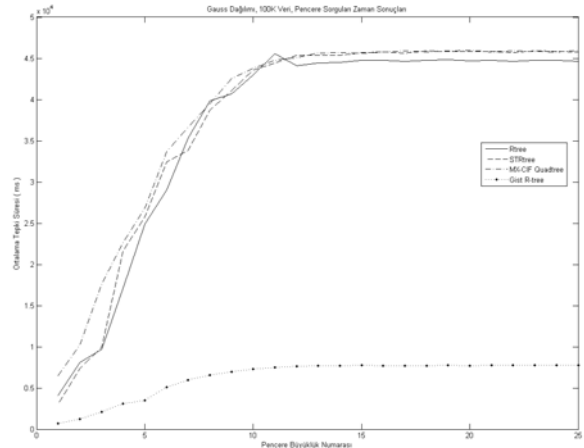
**Tablo 4:** GD\_20K ve GD100K dizin oluşum bilgileri

|               | Gauss Dağılımı, 20K veri |           |                   |                        | Gauss Dağılımı, 100K veri |           |                   |                        |
|---------------|--------------------------|-----------|-------------------|------------------------|---------------------------|-----------|-------------------|------------------------|
|               | R-Ağacı                  | STR-Ağacı | MX-CIF 4'lü Ağacı | Gist R-tree PostgreSQL | R-Ağacı                   | STR-Ağacı | MX-CIF 4'lü Ağacı | Gist R-tree PostgreSQL |
| <b>PBN 1</b>  | 589,06                   | 654,38    | 1004,38           | 175,63                 | 4095,94                   | 3022,81   | 6505              | 691,25                 |
| <b>PBN 2</b>  | 1812,19                  | 1452,81   | 1850,94           | 278,44                 | 8094,69                   | 7336,25   | 10260,63          | 1246,25                |
| <b>PBN 3</b>  | 2722,5                   | 2459,38   | 2650              | 394,69                 | 9692,19                   | 9998,44   | 17602,81          | 2122,19                |
| <b>PBN 4</b>  | 4127,81                  | 3605      | 3732,19           | 521,25                 | 17082,19                  | 21655,63  | 22638,75          | 3095,94                |
| <b>PBN 5</b>  | 4736,25                  | 4150,63   | 4860,63           | 720                    | 24894,06                  | 25745,94  | 26843,13          | 3524,06                |
| <b>PBN 6</b>  | 6266,56                  | 5360,63   | 5894,69           | 906,56                 | 29032,19                  | 32400,94  | 33724,69          | 5115,31                |
| <b>PBN 7</b>  | 6488,44                  | 5996,25   | 6481,88           | 1099,38                | 35278,44                  | 33846,88  | 36610             | 5967,5                 |
| <b>PBN 8</b>  | 7207,81                  | 6880,31   | 7087,81           | 1189,69                | 39860,63                  | 38710     | 39507,19          | 6575,94                |
| <b>PBN 9</b>  | 7714,38                  | 7285,63   | 7589,38           | 1240,94                | 40727,81                  | 41098,75  | 42590,63          | 6965                   |
| <b>PBN 10</b> | 8164,38                  | 7712,81   | 8013,44           | 1337,81                | 42955                     | 43538,75  | 43780             | 7330                   |
| <b>PBN 11</b> | 8387,81                  | 7850,94   | 8123,75           | 1404,06                | 45556,25                  | 44447,5   | 44780,94          | 7483,75                |
| <b>PBN 12</b> | 8426,25                  | 8173,13   | 8318,75           | 1434,38                | 44100,31                  | 45395,94  | 45095,94          | 7600,63                |
| <b>PBN 25</b> | 8711,56                  | 8540,31   | 8698,75           | 1507,81                | 44652,81                  | 45739,69  | 45968,13          | 7739,38                |

**Tablo 5:** GD\_20K ve GD100K, ortalama pencere sorgusu zamanları (ms)



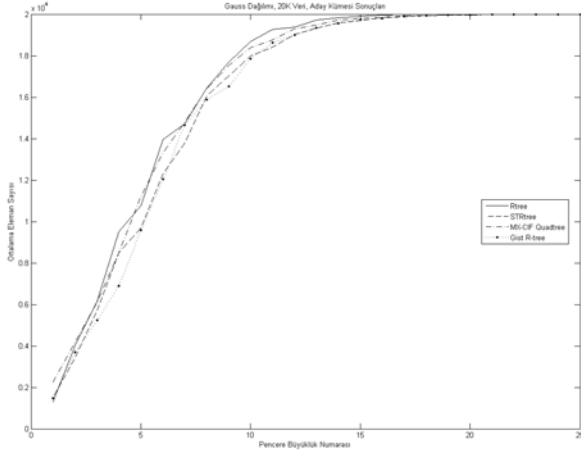
**Şekil 6:** Gauss Dağılımı, 20K Veri, Pencere Sorguları Zaman Sonuçları



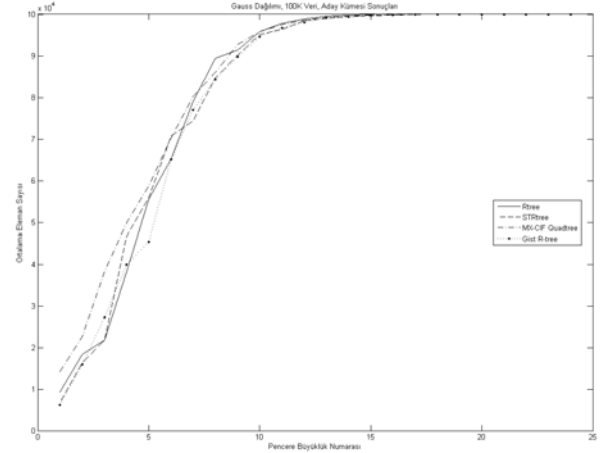
**Şekil 7:** Gauss Dağılımı, 100K Veri, Pencere Sorguları Zaman Sonuçları

|        | Gauss Dağılımı, 20K veri |           |                   |                        | Gauss Dağılımı, 100K veri |           |                   |                        |
|--------|--------------------------|-----------|-------------------|------------------------|---------------------------|-----------|-------------------|------------------------|
|        | R-Ağacı                  | STR-Ağacı | MX-CIF 4'lü Ağacı | Gist R-tree PostgreSQL | R-Ağacı                   | STR-Ağacı | MX-CIF 4'lü Ağacı | Gist R-tree PostgreSQL |
| PBN 1  | 1303,14                  | 1496,76   | 2264,28           | 1491,1                 | 9204,06                   | 6684,74   | 14173,06          | 6138,18                |
| PBN 2  | 4018,56                  | 3429,84   | 4178,5            | 3693,34                | 18279,14                  | 16336,76  | 22541,46          | 15893,78               |
| PBN 3  | 6167,68                  | 5706,88   | 6107,52           | 5247,62                | 21797,02                  | 21861,28  | 38335,12          | 27209,94               |
| PBN 4  | 9493,52                  | 8493,98   | 8491,6            | 6898,74                | 38115,06                  | 46541,82  | 49788,98          | 39875,54               |
| PBN 5  | 10756,38                 | 9680,56   | 11225,16          | 9597,06                | 55536,62                  | 55989,58  | 58776,58          | 45350,04               |
| PBN 6  | 13962,18                 | 12305,42  | 13317,78          | 12050,08               | 65273,58                  | 70690,64  | 70338,82          | 65169,54               |
| PBN 7  | 14699,88                 | 13773,64  | 14786,44          | 14652                  | 79005,88                  | 74267,16  | 80223,68          | 76953,08               |
| PBN 8  | 16437,6                  | 16060,9   | 16420,82          | 15878,76               | 89386,74                  | 84707,22  | 86021,54          | 84373,36               |
| PBN 9  | 17684,46                 | 17020,6   | 17526,32          | 16509,24               | 91460,72                  | 90152,8   | 92774,26          | 89764,18               |
| PBN 10 | 18683,54                 | 17995,36  | 18393,92          | 17850,36               | 95762,94                  | 95085,28  | 95736,06          | 94568,86               |
| PBN 11 | 19277,06                 | 18407,6   | 18764,78          | 18638,68               | 97748,14                  | 96343,98  | 97599,1           | 96775,84               |
| PBN 12 | 19361,26                 | 19036,7   | 19286             | 18992,48               | 98926,92                  | 98519,84  | 98574,34          | 98106,16               |
| PBN 25 | 20000                    | 20000     | 20000             | 20000                  | 100000                    | 100000    | 100000            | 100000                 |

Tablo 6: GD\_20K ve GD100K, ortalama aday küme eleman sayıları



Şekil 8: Gauss Dağılımı, 20K Veri, Aday Kümesi Sonuçları



Şekil 9: Gauss Dağılımı, 100K Veri, Aday Kümesi Sonuçları

Gauss dağılımına sahip her iki veri kümesiyle de yapılan testlerin sonucunda zaman eğrilerinin, sorgu penceresi büyüklük numarası 12 veya 13'e ulaştıktan sonra fazla değişim göstermediği görülmüştür. Test senaryolarına göre pencereler veri uzayının tamamını kapsayacak şekilde büyüdüğünden ve bu dağılıma sahip veriler orta bölgelerde toplandığından, belli bir pencere büyüklüğünden sonra kapsanan veri sayısında önemli artış olmamaktadır. Bu da performans eğrilerinde bir yakınsamaya yol açmaktadır.

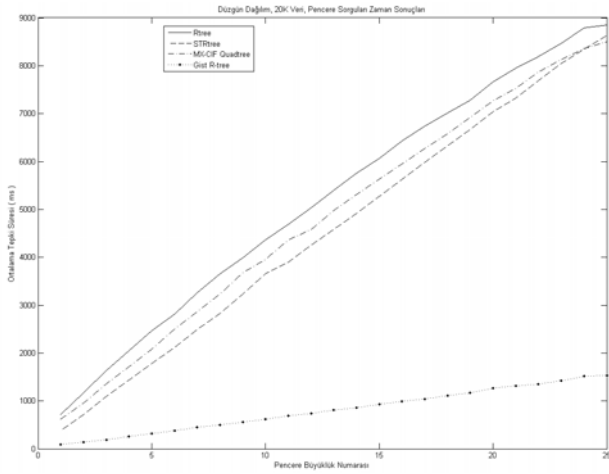
## 5.2 Düzgün Dağılım Sonuçları

|                    | Düzgün Dağılım, 20K veri |           |                  | Düzgün Dağılım, 100K veri |           |                  |
|--------------------|--------------------------|-----------|------------------|---------------------------|-----------|------------------|
|                    | R-Ağacı                  | STR-Ağacı | MXCIF 4'lü Ağacı | R-Ağacı                   | STR-Ağacı | MXCIF 4'lü Ağacı |
| Derinlik           | 5                        | 5         | 22               | 6                         | 6         | 26               |
| Oluşum Zamanı (ms) | 6131,88                  | 6457,50   | 8166,25          | 67136,88                  | 32679,38  | 40973,13         |

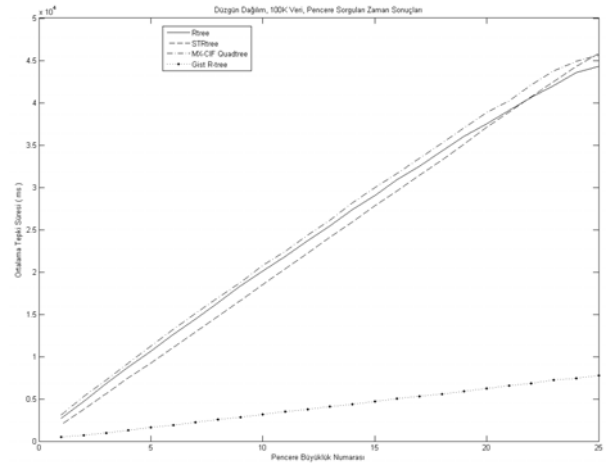
Tablo 7: DD\_20K ve DD100K dizin oluşum bilgileri

|               | Düzgün Dağılım, 20K veri |           |                   |                        | Düzgün Dağılım, 100K veri |           |                   |                        |
|---------------|--------------------------|-----------|-------------------|------------------------|---------------------------|-----------|-------------------|------------------------|
|               | R-Ağacı                  | STR-Ağacı | MX-CIF 4'lü Ağacı | Gist R-tree PostgreSQL | R-Ağacı                   | STR-Ağacı | MX-CIF 4'lü Ağacı | Gist R-tree PostgreSQL |
| <b>PBN 1</b>  | 719,06                   | 381,56    | 612,81            | 86,88                  | 2682,50                   | 1945,31   | 3108,75           | 520,31                 |
| <b>PBN 2</b>  | 1169,06                  | 712,81    | 954,69            | 135,31                 | 4705,31                   | 3767,81   | 5325,94           | 690,00                 |
| <b>PBN 3</b>  | 1636,25                  | 1098,75   | 1357,81           | 190,94                 | 6835,63                   | 5616,56   | 7239,38           | 979,06                 |
| <b>PBN 4</b>  | 2051,88                  | 1428,13   | 1706,88           | 253,75                 | 8770,94                   | 7472,19   | 9234,38           | 1295,63                |
| <b>PBN 5</b>  | 2463,75                  | 1786,25   | 2086,25           | 315,94                 | 10665,63                  | 9259,69   | 11203,44          | 1620,94                |
| <b>PBN 11</b> | 4680,31                  | 3898,75   | 4362,19           | 686,88                 | 21815,94                  | 20375,00  | 22415,00          | 3495,31                |
| <b>PBN 12</b> | 5028,44                  | 4249,06   | 4578,44           | 738,13                 | 23720,00                  | 22234,06  | 24369,69          | 3791,25                |
| <b>PBN 13</b> | 5397,19                  | 4578,75   | 4971,56           | 806,56                 | 25429,38                  | 24080,00  | 26123,13          | 4078,13                |
| <b>PBN 14</b> | 5748,75                  | 4916,56   | 5311,88           | 857,50                 | 27348,75                  | 25927,81  | 28141,88          | 4375,94                |
| <b>PBN 15</b> | 6067,81                  | 5271,88   | 5628,13           | 924,38                 | 29015,63                  | 27819,06  | 29940,94          | 4670,94                |
| <b>PBN 21</b> | 7945,00                  | 7330,31   | 7532,81           | 1314,38                | 39131,88                  | 38894,38  | 40230,31          | 6597,81                |
| <b>PBN 22</b> | 8186,56                  | 7691,25   | 7871,88           | 1352,19                | 40656,56                  | 40701,25  | 42095,94          | 6852,50                |
| <b>PBN 23</b> | 8460,63                  | 8044,38   | 8132,81           | 1419,06                | 42057,19                  | 42477,19  | 43748,44          | 7208,75                |
| <b>PBN 24</b> | 8795,00                  | 8345,00   | 8352,81           | 1513,44                | 43565,63                  | 44329,69  | 44913,75          | 7460,00                |
| <b>PBN 25</b> | 8845,00                  | 8639,69   | 8503,75           | 1528,44                | 44303,44                  | 45864,69  | 45635,63          | 7790,31                |

**Tablo 8:** DD\_20K ve DD100K, ortalama pencere sorgusu zamanları (ms)



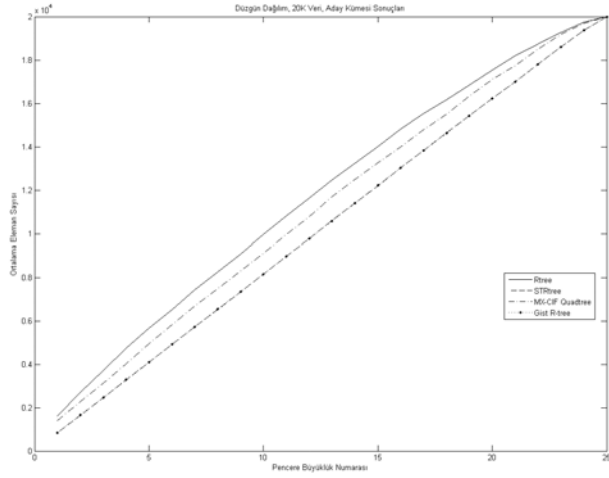
**Şekil 10:** Düzgün Dağılım, 20K Veri, Pencere Sorguları Zaman Sonuçları



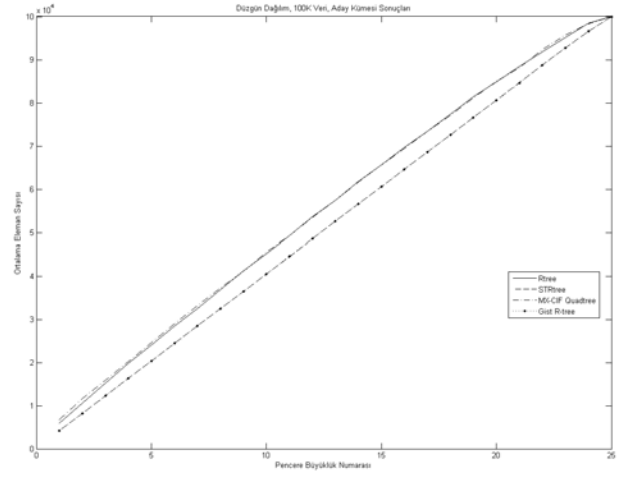
**Şekil 11:** Düzgün Dağılım, 100K Veri, Pencere Sorguları Zaman Sonuçları

|        | Düzgün Dağılım, 20K veri |           |                   |                        | Düzgün Dağılım, 100K veri |           |                   |                        |
|--------|--------------------------|-----------|-------------------|------------------------|---------------------------|-----------|-------------------|------------------------|
|        | R-Ağacı                  | STR-Ağacı | MX-CIF 4'lü Ağacı | Gist R-tree PostgreSQL | R-Ağacı                   | STR-Ağacı | MX-CIF 4'lü Ağacı | Gist R-tree PostgreSQL |
| PBN 1  | 1615,18                  | 844,44    | 1392,94           | 839,82                 | 5984,52                   | 4157,30   | 6790,22           | 4155,32                |
| PBN 2  | 2717,36                  | 1652,56   | 2273,50           | 1656,58                | 10567,84                  | 8226,62   | 11631,24          | 8223,44                |
| PBN 3  | 3720,40                  | 2477,82   | 3142,82           | 2474,80                | 15333,28                  | 12271,46  | 15907,14          | 12278,42               |
| PBN 4  | 4728,18                  | 3276,94   | 4013,76           | 3290,34                | 19755,68                  | 16332,76  | 20222,50          | 16337,02               |
| PBN 5  | 5656,92                  | 4087,30   | 4938,38           | 4102,12                | 24004,72                  | 20363,36  | 24590,04          | 20356,72               |
| PBN 11 | 10807,00                 | 8966,80   | 9967,30           | 8985,78                | 49329,74                  | 44512,50  | 49205,90          | 44548,12               |
| PBN 12 | 11626,98                 | 9764,04   | 10783,32          | 9774,96                | 53650,76                  | 48559,70  | 53468,20          | 48559,34               |
| PBN 13 | 12469,42                 | 10596,76  | 11687,50          | 10586,38               | 57342,92                  | 52568,38  | 57340,68          | 52569,08               |
| PBN 14 | 13250,08                 | 11388,54  | 12502,48          | 11409,48               | 61850,80                  | 56589,02  | 61674,80          | 56589,24               |
| PBN 15 | 14003,46                 | 12212,72  | 13269,32          | 12232,52               | 65682,78                  | 60580,88  | 65601,20          | 60586,66               |
| PBN 21 | 18190,82                 | 17008,68  | 17753,56          | 17003,68               | 88454,68                  | 84700,74  | 88238,40          | 84641,52               |
| PBN 22 | 18746,14                 | 17799,62  | 18493,04          | 17795,42               | 91826,52                  | 88704,86  | 92284,84          | 88689,98               |
| PBN 23 | 19276,24                 | 18595,06  | 19176,66          | 18594,74               | 95158,22                  | 92712,76  | 95711,38          | 92689,66               |
| PBN 24 | 19736,78                 | 19364,40  | 19693,76          | 19366,00               | 98456,00                  | 96662,22  | 98328,56          | 96663,56               |
| PBN 25 | 20000,00                 | 19997,24  | 19999,44          | 19997,30               | 100000,00                 | 99988,34  | 99997,38          | 99988,92               |

Tablo 9: DD\_20K ve DD100K, ortalama aday küme eleman sayıları



Şekil 12: Düzgün Dağılım, 20K Veri, Aday Kümesi Sonuçları



Şekil 13: Düzgün Dağılım, 100K Veri, Aday Kümesi Sonuçları

Düzgün dağılıma sahip veri kümeleriyle elde edilen eğrilerde herhangi bir yakınsama olmamıştır. Sorgu pencereleri her adımda büyüdüğünden ve nereye yerleşirse yerleşsin veri uzayının her tarafına dağılmış olan veriler ile örtüştüğünden dolayı sürekli olarak artış gösteren sonuçlar alınmıştır. Pencereler tüm uzayı kapsayacak büyüklüğe ulaştığında ise yöntemler denk duruma gelmektedir.

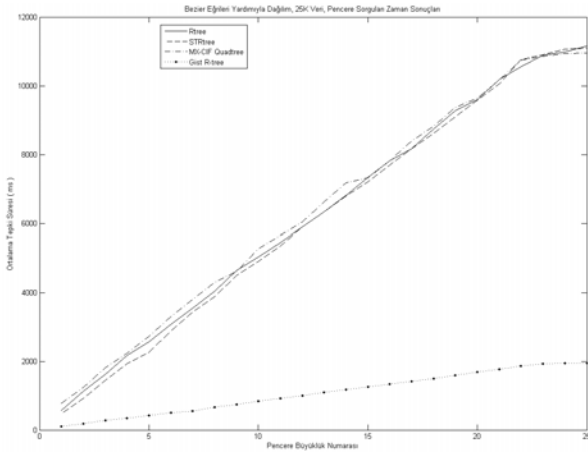
### 5.3 Bezier Eğrileri ile Dağılım Sonuçları

|                    | Bezier Eğrileri ile Dağılım, 25K veri |           |                  |
|--------------------|---------------------------------------|-----------|------------------|
|                    | R-Ağacı                               | STR-Ağacı | MXCIF 4'lü Ağacı |
| Derinlik           | 6                                     | 5         | 16               |
| Oluşum Zamanı (ms) | 7983,75                               | 8226,88   | 8742,50          |

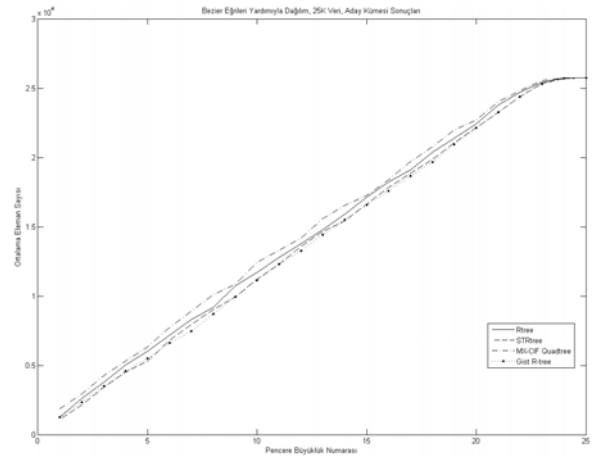
Tablo 10: BeU\_25K dizin oluşum bilgileri

|               | Bezier Eğrileri ile Dağılım, 25K veri |           |                   |                        |                                  |           |                   |                        |
|---------------|---------------------------------------|-----------|-------------------|------------------------|----------------------------------|-----------|-------------------|------------------------|
|               | Pencere Sorgu Zamanları (ort. ms)     |           |                   |                        | Aday Küme Eleman Sayıları (ort.) |           |                   |                        |
|               | R-Ağacı                               | STR-Ağacı | MX-CIF 4'lü Ağacı | Gist R-tree PostgreSQL | R-Ağacı                          | STR-Ağacı | MX-CIF 4'lü Ağacı | Gist R-tree PostgreSQL |
| <b>PBN 1</b>  | 561,56                                | 467,5     | 781,56            | 103,13                 | 561,56                           | 467,5     | 781,56            | 103,13                 |
| <b>PBN 2</b>  | 1120,94                               | 913,75    | 1241,25           | 180,63                 | 1120,94                          | 913,75    | 1241,25           | 180,63                 |
| <b>PBN 3</b>  | 1619,38                               | 1454,38   | 1823,13           | 275,31                 | 1619,38                          | 1454,38   | 1823,13           | 275,31                 |
| <b>PBN 4</b>  | 2175,94                               | 1922,19   | 2234,38           | 348,13                 | 2175,94                          | 1922,19   | 2234,38           | 348,13                 |
| <b>PBN 5</b>  | 2560,63                               | 2259,69   | 2722,81           | 419,69                 | 2560,63                          | 2259,69   | 2722,81           | 419,69                 |
| <b>PBN 11</b> | 5447,81                               | 5355,63   | 5660,63           | 925                    | 5447,81                          | 5355,63   | 5660,63           | 925                    |
| <b>PBN 12</b> | 5898,44                               | 5894,06   | 6038,44           | 999,06                 | 5898,44                          | 5894,06   | 6038,44           | 999,06                 |
| <b>PBN 13</b> | 6326,25                               | 6330,63   | 6620,94           | 1092,19                | 6326,25                          | 6330,63   | 6620,94           | 1092,19                |
| <b>PBN 14</b> | 6803,75                               | 6785      | 7168,75           | 1175,31                | 6803,75                          | 6785      | 7168,75           | 1175,31                |
| <b>PBN 15</b> | 7342,81                               | 7205,94   | 7327,5            | 1254,38                | 7342,81                          | 7205,94   | 7327,5            | 1254,38                |
| <b>PBN 21</b> | 10177,81                              | 10062,5   | 10174,69          | 1766,88                | 10177,81                         | 10062,5   | 10174,69          | 1766,88                |
| <b>PBN 22</b> | 10558,75                              | 10766,56  | 10745             | 1868,75                | 10558,75                         | 10766,56  | 10745             | 1868,75                |
| <b>PBN 23</b> | 10893,75                              | 10905,94  | 10850,63          | 1926,56                | 10893,75                         | 10905,94  | 10850,63          | 1926,56                |
| <b>PBN 24</b> | 10990,63                              | 11058,13  | 10940,63          | 1944,69                | 10990,63                         | 11058,13  | 10940,63          | 1944,69                |
| <b>PBN 25</b> | 8845                                  | 8639,69   | 8503,75           | 1528,44                | 8845                             | 8639,69   | 8503,75           | 1528,44                |

Tablo 11: BeU25K, ortalama pencere sorgusu zamanları (ms) ve ortalama aday küme eleman sayıları



Şekil 14: Bezier Eğrileri ile Dağılım, 25K Veri, Pencere Sorguları Zaman Sonuçları



Şekil 15: Bezier Eğrileri ile Dağılım, 25K Veri, Aday Kümesi Sonuçları

## 6. DEĞERLENDİRME

Konumsal dizinleme yöntemlerinin başarımlarını karşılaştırmak amacıyla yapılan bu çalışmada R-ağacı, STR-ağacı, MX-CIF 4'lü ağacı ve VTYS ile bütünleşik olan GiST R-ağacı kullanılmıştır. İstatistiksel dağılımlar ve Bezier eğrileri yardımıyla üretilen sentetik veri kümeleri kullanılarak uygulanan test senaryolarıyla da pencere sorgusu başarımları ölçülmüş ve karşılaştırılmıştır.

İlk olarak VTYS üzerinde dizinleme katmanı kurularak gerçekleştirilen dizin yapılarının sonuçlarını ele alalım. Sonuç tabloları ve grafikleri beraber incelendiğinde üç yöntem arasında sorgu başarımı en yüksek olan STR-ağacıdır. Ön işlemlerle düzenlenen veriler sayesinde ağacın yaprakları daha iyi gruplanmakta ve doluluk oranları yükseltilmektedir. Ağaç daha düzenli bir yapıya kavuştuğundan dolayı hem sorgu zamanları hem de aday küme eleman sayıları açısından diğer iki yöntemi geride bırakmayı başarmaktadır. Fakat sorgu başarımı konusunda öne çıkan bu yöntemin içsel yapısı ve tanımı gereği bir dezavantajı vardır. Sık değişim göstermeyen, durağan ortamlarda etkili sorgu başarımı gösterebilecek olan STR-ağacı, ekleme ve silme işlemlerinin bol olduğu dinamik ortamlarda yeniden yapılandırmaya ihtiyaç duyacaktır. Bu yönüyle düşünüldüğünde verilerin sürekli güncellenmediği veya

güncellense bile belirli aralıklarla bakım yapma, yeniden yapılandırma ihtiyacının rahatlıkla karşılanabileceği alanlarda tercih edilmesi daha uygun olacaktır. R-ağacı ve MX-CIF 4'lü ağacı dinamik ortamlara yapısal olarak daha uygundur. Veriler dizinlenmeden önce ön işlemlerin uygulanmadığı bu yöntemler ekleme ve silme işlemleri sonrasında yapılarını dinamik olarak düzenleyebilirler. Sorgu başarımları açısından incelendiklerinde ise kendi aralarındaki fark, STR-ağacı ile aralarındaki kadar açık değildir. Deney düzeneğinde belirtildiği üzere kullanılan R-ağacının dallanma faktörü 10'dur. MX-CIF 4'lü ağacının ise tanımı gereği 4'tür. Veri artışıdaki yükseliş, söz konusu iki ağacın derinlikleri arasındaki farkın bir anda açılmasına sebep olabilmektedir. Sonuçları bir de ayrık katmanı ve bütünleşik mimari açısından ele alırsak bir başka çarpıcı durum ortaya çıkmaktadır. Süzme işlemi her yöntem ve her iki mimari yaklaşımda da başarılı tamamlanabilmektedir.

## KAYNAKLAR

**Kedem G.**, 1982, *The Quad-CIF tree: a data structure for hierarchical on-line algorithms*, Proceedings of the Nineteenth Design Automation Conference, January, Las Vegas, United States p.352-357

**Guttman A.**, 1984, *R-Trees: A Dynamic Index Structure for Spatial Searching*, In Proceedings of the ACM SIGMOD Conference, June, Boston, United States, p 47-57

**Kamel I., Faloutsos C.**, 1993, *On Packing R-trees*, Proceedings of 2nd International Conference on Information and Knowledge Management (CKIM-93), November , Arlington, VA, p. 490-499

**Leutenegger S.T., Lopez M.A., Edgington J.**, 1997, *STR: A Simple and Efficient Algorithm for R-tree Packing*, Proceedings of the 13th International Conference on Data Engineering (ICDE 1997), 7-11 April, Birmingham, UK, p 497-506

**RAPOR 1**, *PostGIS Manual, Working Paper*, 30.12.2005

**Stefanakis E., Sellis T.**, 1997, *Towards The Design of A DBMS Repository for The Application Domain of GIS, Requirements of Users and Applications*, Proceedings of the 18th International Cartographic Conference, 21-27 June, Stockholm, Sweden, p 2030-2037

**URL 1**, Wikipedia Free Encyclopedia, *Bézier Curve*, [http://en.wikipedia.org/wiki/B%C3%A9zier\\_curve](http://en.wikipedia.org/wiki/B%C3%A9zier_curve), Son erişim tarihi: 31.01.2007

**Yomrahoğlu T.**, 2002. *Coğrafi Bilgi Sistemleri: Temel Kavramlar ve Uygulamalar*, Akademi Kitabevi, Trabzon, 36 s.