ÖZET

VERİ YAPISININ VE İÇERİĞİNİN ANALİZ EDİLMESİYLE BENZER VERİLERİN BELİRLENMESİNİ SAĞLAYAN BİR SİSTEM

Bu buluş, verilerin depolandığı kaynaklara erişerek verinin verisinin (metadata - metaveri) işlenmesini ve bu verilerin içerikleri ile yapılarının analiz edilmesi yöntemi ile benzer verilerin otomatik olarak tespit edilmesini sağlayan bir sistem (1) ile ilgilidir.

İSTEMLER

1. Verilerin depolandığı kaynaklara erişerek verinin verisinin (metadata - metaveri) işlenmesini ve bu verilerin içerikleri ile yapılarının analiz edilmesi yöntemi ile benzer verilerin otomatik olarak tespit edilmesini sağlayan;

- üzerinde en az bir uygulama çalıştırmak üzere yapılandırılan en az bir elektronik cihaz (2) içeren ve
- elektronik cihaz (2) üzerinde çalışmak ve kullanıcının tarama ve karşılaştırma yapmak istediği veri tablolarını girmesini mümkün kılmak üzere yapılandırılan en az bir uygulama (3),
- taranacak ve karşılaştırılacak verileri tablolar içerisinde depolamak üzere yapılandırılan en az bir veri tabanı (4),
- kullanıcının uygulama (3) üzerinden sağladığı talimat doğrultusunda taranması istenen veri tablosuna veri tabanı (4) üzerinden erişmek, tarama yapacağı tablodaki veri tiplerini belirlemek, veri dağılım istatistiklerini analiz etmek, tarama yapacağı tablodaki kolonlarda bulunan en yüksek ve en düşük değerleri belirlemek, tarama yapacağı verileri üzerinde veri dağılımlarının normalizasyonlarını ve standart dağılımlarını hesaplayarak yorumlamak, tarama işlemini tamamladıktan sonra istatistiksel olarak kolonların benzerliğini kıyaslamak veverilerin doğruluk yüzdesine göre karşılaştırma yöntemi kullanarak kolonların veri içeriğine göre aynı veriyi içerip içermediklerini belirlemek üzere yapılandırılan en az bir sunucu (5) ile karakterize edilen bir sistem (1).

2. Üzerinde en az bir uygulama çalıştırmak üzere yapılandırılan elektronik cihaz (2) ile karakterize edilen İstem 1'deki gibi bir sistem (1).

3. Elektronik cihaz (2) üzerinde çalışmak ve kullanıcının taranmasını ve karşılaştırılmasını istediği veri tablolarına ilişkin bilgi girişi yapmasını

mümkün kılmak üzere yapılandırılan uygulama (3) ile karakterize edilen İstem 1 veya 2'deki gibi bir sistem (1).

4. Şirkete ilişkin verileri ve verilere ilişkin verileri depolamak üzere yapılandırılan veri tabanı (4) ile karakterize edilen yukarıdaki istemlerden herhangi birindeki gibi bir sistem (1).

5. Kullanıcının uygulama (3) üzerinden girdiği bilgiler doğrultusunda taranacak olan verileri veri tabanına (4) erişerek almak üzere yapılandırılan sunucu (5) ile karakterize edilen yukarıdaki istemlerden herhangi birindeki gibi bir sistem (1).

6. Kullanıcının uygulama (3) üzerinden girdiği bilgiler doğrultusunda karşılaştırılacak olan verileri veri tabanına (4) erişerek almak üzere yapılandırılan sunucu (5) ile karakterize edilen yukarıdaki istemlerden herhangi birindeki gibi bir sistem (1).

7. Tarama yapacağı tablodaki kolonların veri dağılımlarına ilişkin istatistiksel hesaplamalar yapmak üzere yapılandırılan sunucu (5) ile karakterize edilen yukarıdaki istemlerden herhangi birindeki gibi bir sistem (1).

TARİFNAME

VERİ YAPISININ VE İÇERİĞİNİN ANALİZ EDİLMESİYLE BENZER VERİLERİN BELİRLENMESİNİ SAĞLAYAN BİR SİSTEM

Teknik Alan

Bu buluş, verilerin depolandığı kaynaklara erişerek verinin verisinin (metadata - metaveri) işlenmesini ve bu verilerin içerikleri ile yapılarının analiz edilmesi yöntemi ile benzer verilerin otomatik olarak tespit edilmesini sağlayan bir sistem ile ilgilidir.

Önceki Teknik

Günümüzde mevcut veri modelleri yönetimi farklı ticari modelleme ürünleri üzerinden yapılmaktadır. Bu ürünlerin genel özelliği modeli ürün üzerinde oluşturma ve veri tabanları ile ürün üzerinden yönetimin yapılmasını sağlamaktır. Ürünlerde farklı kaynakları aynı ortama alarak karşılaştırma imkanı sunulmaktadır. Ancak veri sayısı arttıkça verilerin depolandığı konumlar birbirinden uzaklaşmakta ve verilerin birbirleri ile kıyaslanması için gereken efor artmaktadır.

Mevcut teknikte bulunan çalışmalar göz önünde bulundurulduğunda büyük ve karmaşık veri entegrasyonlarına sahip firmaların verilerin yapısını ve içeriğini analiz ederek nerelerde benzer veriler bulunduğunu tespit eden bir sisteme ihtiyaç duyulduğu anlaşılmaktadır.

Tekniğin bilinen durumunda yer alan US2010005048 sayılı Birleşik Devletler patent dokümanında, veri yönetimine ve daha özel olarak çift kayıtların saptanması ve çıkarılmasını içeren bir sistemden bahsedilmektedir. Söz konusu sistem; ikili kayıtların saptanması ve çıkarılması için bir yöntem temin etmektedir. Burada her kayıt, çok sayıda alana ilişkin verilerden oluşmaktadır. Yöntem, alana özgü bilgi

tabanını kullanarak verilerin standartlaştırılması adımlarını içermektedir. Bir veya daha fazla ilgili kayıt alanlarının en azından bir kısmının çıkarılması, "karşılaştırılabilir" alan kısmına ayıklanan veriler üzerinde anahtarlar oluşturmak için eşleşen bir özellik işlevi uygulamak, oluşturulan alan seviyesi tuşlarını kullanarak kayıt seviyesi anahtarları oluşturmak, oluşturulan kayıt seviyesi tuşlarına göre kayıtların kümelenmesi, tanımlanan her bir küme için referans kaydının belirlenmesi ve kümenin referans kaydına göre bir kümedeki her kayıt için eşleşme yüzdesini hesaplamayı içermektedir.

Buluşun Kısa Açıklaması

Bu buluşun amacı, büyük ve karmaşık veri entegrasyonlarına sahip şirketler için, sahip oldukları verileri analiz ederek benzer verilerin nerede depolandığının belirlenmesini sağlayan bir sistem gerçekleştirmektir.

Buluşun Ayrıntılı Açıklaması

Bu buluşun amacına ulaşmak için gerçekleştirilen "Veri Yapısının Ve İçeriğinin Analiz Edilmesiyle Benzer Verilerin Belirlenmesini Sağlayan Bir Sistem" ekli şekilde gösterilmiş olup, bu şekil;

Şekil-1; Buluş konusu sistemin şematik bir görünüşüdür.

Şekillerde yer alan parçalar tek tek numaralandırılmış olup, bu numaraların karşılıkları aşağıda verilmiştir.

1. Sistem
2. Elektronik cihaz
3. Uygulama
4. Veri tabanı
5. Sunucu

Buluş konusu, verilerin depolandığı kaynaklara erişerek verinin verisinin (metadata - metaveri) işlenmesini ve bu verilerin içerikleri ile yapılarının analiz edilmesi yöntemi ile benzer verilerin otomatik olarak tespit edilmesini sağlayan bir sistem (1);

- üzerinde en az bir uygulama çalıştırmak üzere yapılandırılan en az bir elektronik cihaz (2),
- elektronik cihaz (2) üzerinde çalışmak ve kullanıcının tarama ve karşılaştırma yapmak istediği veri tablolarını girmesini mümkün kılmak üzere yapılandırılan en az bir uygulama (3),
- taranacak ve karşılaştırılacak verileri tablolar içerisinde depolamak üzere yapılandırılan en az bir veri tabanı (4),
- kullanıcının uygulama (3) üzerinden sağladığı talimat doğrultusunda taranması istenen veri tablosuna veri tabanı (4) üzerinden erişmek ve karşılaştırılması istenen veri tablolarına erişerek veri dağılım istatistiklerini analiz etmek üzere yapılandırılan en az bir sunucu (5) içermektedir.

Buluş konusu sistemde (1) yer alan elektronik cihaz (2) üzerinde en az bir uygulama çalıştırmak üzere yapılandırılmaktadır.

Buluş konusu sistemde (1) yer alan uygulama (3) elektronik cihaz (2) üzerinde çalışmak ve kullanıcının taranmasını ve karşılaştırılmasını istediği veri tablolarına ilişkin bilgi girişi yapmasını mümkün kılmak üzere yapılandırılmaktadır.

Buluş konusu sistemde (1) yer alan veri tabanı (4) şirkete ilişkin verileri ve verilere ilişkin verileri depolamak üzere yapılandırılmaktadır.

Buluş konusu sistemde (1) yer alan sunucu (5) kullanıcının uygulama (3) üzerinden girdiği bilgiler doğrultusunda taranacak olan verileri veri tabanına (4) erişerek almak üzere yapılandırılmaktadır. Sunucu (5) kullanıcının uygulama (3) üzerinden girdiği bilgiler doğrultusunda karşılaştırılacak olan verileri veri tabanına (4)

erişerek almak üzere yapılandırılmaktadır. Sunucu (5) tarama yapacağı tablodaki veri tiplerini belirlemek üzere yapılandırılmaktadır. Sunucu (5) tarama yapacağı tablodaki kolonların veri dağılımlarına ilişkin istatistiksel hesaplamalar yapmak üzere yapılandırılmaktadır. Sunucu (5) tarama yapacağı tablodaki kolonlarda bulunan en yüksek ve en düşük değerleri belirlemek üzere yapılandırılmaktadır. Sunucu (5) tarama yapacağı verileri üzerinde veri dağılımlarının normalizasyonlarını ve standart dağılımlarını hesaplayarak yorumlamak üzere yapılandırılmaktadır. Sunucu (5) tarama işlemini tamamladıktan sonra istatistiksel olarak kolonların benzerliğini kıyaslamak üzere yapılandırılmaktadır. Sunucu (5) verilerin doğruluk yüzdesine göre karşılaştırma yöntemi kullanarak kolonların veri içeriğine göre aynı veriyi içerip içermediklerini belirlemek üzere yapılandırılmaktadır.

Buluş konusu sistemde (1) öncelikle kullanıcı, elektronik cihaz (2) üzerinde çalışan uygulama (3) üzerinden taranmasını ve karşılaştırılmasını istediği verilere ilişkin bilgiler girmektedir. Bir veri tabanında (4) tutulan veriler, sunucu (5) tarafından kullanıcının girdiği bilgiler doğrultusunda alınmaktadır. Daha sonra sunucu (5) verileri istatistiksel olarak analiz ederek normalizasyonları ve standart dağılımları incelemekte ve yorumlamaktadır. Sunucu (5) yaptığı hesaplamalar ve analizler sonucunda kolonlarda yer alan verilerin aynı olup olmadığını belirlemek üzere yapılandırılmaktadır.

Bu temel kavramlar etrafında, buluş konusu sistem (1) ile ilgili çok çeşitli uygulamaların geliştirilmesi mümkün olup, buluş burada açıklanan örneklerle sınırlandırılamaz, esas olarak istemlerde belirtildiği gibidir.

Şekil 1

