

ÜÇÜNCÜ NORMAL FORM KULLANILARAK VERİ ANALİZİ TEKNİĞİ

Ramazan YANAR

ÖZET

Mevcut sorunları giderecek ve yeni gereksinimleri karşılayacak bir veri yapısı tasarımı veri tabanının en belirleyici özelliğidir. Bu makalede varlıklara ait grafik olmayan bilgilerin manyetik ortamda (Örneğin: disk, teyp, vb. üzerinde) sunulmasına ilişkin analiz ve tasarım, Veri Analizi Tekniği kullanılarak, bir örnek üzerinde açıklanmaktadır.

ABSTRACT

Designing of a data structure which will solve present problems and meet the new requirements is the most characteristic feature of the database. In this paper analysis and design of the representation of the non-graphic data, in the magnetic medium (i.e. on disks, tapes, etc.) for the entities will be examined upon an example using the Third Normal Form Data Analysis Technique.

1. GİRİŞ

Veri Tabanı'ndaki (VT) grafik olmayan bilgiler bir Veri Tabanı Yönetim Sistemi (VTYS) ile yönetilecektir. Bu VTYS'leri genellikle hiyerarşik, ağ ve ilişkisel veri modellerinden birine yöneliktirler. Veri analizinde ilişkisel model yaklaşımı çok büyük yaygınlık kazanmıştır.

Bugünkü veri tabanı sistemi yazılımlarının pek çoğu mevcut bir ilişkisel VTYS ile (Örneğin, INFO, INFORMIX, ORACLE vb.) bütünleştirilmiştir. Grafik olmayan verileri ilişkisel bir VTYS'nde organize etmek için üçüncü Normal Form (ÜNF) Veri Analizi Tekniği (Normalleştirme) işleminden geçirilmesi gerekmektedir.

Normalleştirmede; önce öznitelikler belirlenir. Daha sonra birinci, ikinci ve üçüncü normal formlar hazırlanarak bunlar optimize edilir. Böylece sonuçta ilişkisel veri tabanında bulunacak tablolar elde edilmiş olur /2/, /4/.

2. NORMALLEŞTİRME

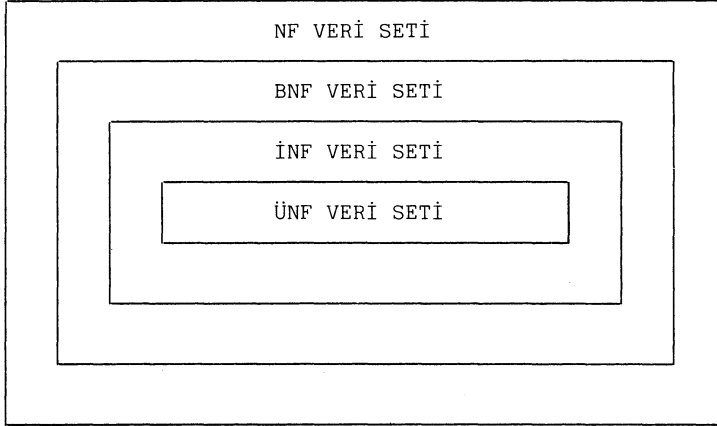
Normalleştirme teorisi, normal formlar kavramına dayalı bir teoridir (Şekil-1). İçerisinde tekrarlı veri grupları ve veri fazlalığı olan klasik bilgi formlarının daha anlaşılır hale getirilmesi ve ihtiva ettikleri bilgiye erişimi kolaylaştıracak veri setlerinin mantıksal tasarımı hakkında yol gösterici bir tekniktir /1/. Normalleştirme adımlarını aşağıdaki gibi sıralamak olasıdır.

* Varlıklar ve bunlara ait öznitelikler sıralanır. Buna Normalleştirilmemiş Form denir.

* NF, birinci normal forma (BNF) çevrilir. Birinci normal formda,tekrarlanan gruplar ayrı veri setleri (ilişkiler) halinde gösterilir.

* BNF, ikinci normal forma (İNF) çevrilir. İNF'da, veri elemanlarının, birleşik anahtarların tümüne ait olup olmadığı tespit edilir.

* İNF, üçüncü normal forma (ÜNF) çevrilir. ÜNF'da, anahtar dışı alanlardaki fonksiyonel bağımlılıklar tespit edilir. Normalleştirmenin bu adımları, anlatımı kolaylaştırmak üzere basit bir örnek üzerinde açıklanacaktır /4/.



Şekil-1: Normal Form Veri Setleri

a. Normalleştirilmemiş Form

Tablo-1, personelin bir projeye tahsisini ve Tablo-2'de de, bu personel tahsisatının normalleştirilmemiş bir formunu göstermektedir.

Normalleştirilmemiş (Tablo-2) form, incelendiğinde aşağıdaki sonuçları çıkarmak olasıdır /3/.

- (1) Satırlar (kayıtlar) değişik uzunluktadırlar.
- (2) Personel Numarası ile Tahsis süresi arasındaki sütünlara ait veriler birden çok kere tekrarlanmıştır. Çünkü personele ait bilgiler birbiri ardınca birer satıra yazılmıştır.

b. Birinci Normal Form (BNF)

İlk adımda normalleştirilmemiş formdan tekrarlamalı gruplar ayrılır. Bunu yapabilmek için, ortamdan öncelikle normalleştirilmemiş veri seti için bir anahtar seçilir. Anahtarın seçimi için şu öneriler gözönüne alınır;

Tablo-1: Personel Proje Tahsis Formu

PROJE KODU : VT01		PROJE TİPİ : VERİ TABANI			
TANITIM : Veri tabanı tasarımı projesi					
Personel No.	İsim	Derece	Ücret Katsayısı	Katılım Tarihi	Tahsis Süresi (ay)
2246	Tülay	A1	4	1.3.94	24
3245	Semih	A2	4	2.4.94	24
6226	Yavuz	B1	9	3.4.94	28
2224	Biröl	A2	4	4.4.94	28
8292	Burak	A1	4	1.5.94	22

Tablo-2: Normalleştirilmemiş Form

Proje Kodu	Proje Tipi	Tanıtım	Per.No.	İsim	Derece	Ücr. Kts.	Katılım Tarihi	Tahsis Süresi
VT01	VT	VT tasarımı	2246	Tülay	A1	4	1.3.94	24
			3245	Semih	A2	4	2.4.94	24
			6226	Yavuz	B1	9	3.4.94	28
			2224	Biröl	A2	4	4.4.94	28
			8292	Burak	A1	4	1.5.94	22
BAK04	Bakım	Bakım Ödenek-leri	6242	Yaşar	A2	4	1.5.94	6
			3269	Cem	B2	9	1.6.94	6
			6245	Ali	B3	9	6.7.94	7

- (1) Seçilen anahtar, tablonun bütün satırlarında tek bir değere sahip olmalıdır. Dolayısıyla anahtar tek bir satırı niteler.
- (2) Anahtar, bir satırda birden fazla tekrarlanmamalıdır.
- (3) Anahtar seçiminde, çok sayıda satırı olan bir kolonla, az sayıda satırı olan başka bir kolon arasında seçim yapmak gerektiğinde az sayıda satırı olan kolon anahtar olarak seçilmelidir.
- (4) Mümkünse metne dayanmayan anahtarlar seçilmelidir.

Bir tekrarlamalı grup şöyle ifade edilir; eğer anahtarın tek bir değerine, çok sayıda değer karşılık geliyorsa, bu veri sahası veya veri sahasının bir grubu tekrarlamalı gruptur /3/.

Tablo-1'de, anahtar olarak seçilen "Proje Kodu"nın bir değerine karşılık çok ve değişik sayıda "Personel No., İsim, Derece, Ücret Katsayısı, Katılım Tarihi ve Tahsis Süresi" elde edilmektedir. Bundan dolayı bu kolonlar, verilen bir anahtarın tekrarlamalı gruplarını oluştururlar. Tekrarlanmayan sahalara, Tablo-3'de görüldüğü gibi BNF veri setini oluşturmak üzere ayrılırlar. Tekrarlanmayan veri gruplarının anahtarı hâlâ proje kodudur.

Tablo-3: Tekrarlanmayan Veri Grubu (BNF Veri Seti (No.:1))

PROJE KODU	PROJE TİPİ	TANITIM
VT01	VT	VT Tasarımı
BAK04	Bakım	Bakım Ödenekleri

Tekrarlanan veri ayrı bir veri seti oluşturmak üzere tablodan ayrılır. Bu tekrarlanan grubun anahtarı, anahtar seçiminde gözönünde bulunacak önerilere göre yapılır. Proje kodu (Tablo-3) ve tekrarlanan sahadan Personel Numarası bir birleşik anahtar oluşturmaktadır (Tablo-4). BNF işleminin sonucu Tablo-3 ve Tablo-4'de görülen iki veri setidir.

Tablo-4: Tekrarlanan Veri Grubu (BNF Veri Seti (No.:2))

PROJE KODU	PER.NO.	İSİM	DERECE	ÜCR.KAT.	KAT.TA.	TAH.SÜ.
VT01	2246	Tülay	A1	4	1.3.94	24
VT01	3245	Semih	A2	4	2.4.94	24
VT01	6226	Yavuz	B1	9	3.4.94	28
VT01	2224	Bırol	A2	4	4.4.94	28
VT01	8292	Burak	A1	4	1.5.94	22
BAK04	6242	Yaşar	A2	4	1.5.94	6
BAK04	3269	Cem	B2	9	1.6.94	6
BAK04	6245	Ali	B3	9	6.7.94	7

c. İkinci Normal Form (İNF)

BNF veri setleri İNF'a dönüştürülürken ikinci adımda;

"KİSMİ - ANAHTAR BAĞIMLILIKLARI AYRILIR"

Bu işlem bileşik anahtarı olan veri setlerine uygulanır ve her saha için (anahtar sahalar hariç) şu soru sorulur;

* "Saha anahtarın bir kısmına mı bağlıdır, yoksa bileşik anahtarla ilişkili midir?".

Bu soru BNF veri setine (Tablo-4) sorulduğunda şu sonuçları görmek olasıdır. Proje Kodu ve Personel Numarası bileşik anahtarlarına karşılık;

- (1) İsim, bütün anahtara bağımlı mıdır?
Hayır. Sadece Personel No.'ya bağlıdır.
- (2) Derece, bütün anahtara bağlı mıdır?
Hayır. Sadece Personel No.'ya bağlıdır.
- (3) Ücret Katsayısı, bütün anahtara bağlı mıdır?
Hayır. Sadece Personel No.'ya bağlıdır.
- (4) Katılım Tarihi, bütün anahtara bağlı mıdır?
Evet. Belli bir projeye ilişkin personel ile ilgili bir veridir.
- (5) Tahsis süresi, bütün anahtara bağlı mıdır?
Evet. Belli bir kişinin belli bir projeye tahsisinde, ayrılan zaman süresidir.

Görüldüğü gibi ilk üç saha bileşik anahtarın sadece Personel Numarasına bağlıdır. BNF veri setinde buna benzer anahtar bağımlılığı bulunduğu anda, bağımlı sahalar veri setinden ayrılır ve bağlı olduğu anahtar, anahtarı olacak şekilde ayrı bir veri seti oluşturulur. Tablo-5'te "Personel Numarası"nın anahtar olduğu İNF veri seti görülmektedir. BNF Veri Seti (No.:1) aynı zamanda İNF Veri Seti (No.1) olarak kabul edilirken Tablo-5 ve 6'da, BNF Veri Seti (No.:1)'e ilave olarak, elde edilen İNF Veri Setleri görülmektedir /3/.

Tablo-5: İkinci Normal Form Veri Seti (No.:2)

PER.NO.	İSİM	DERECE	ÜCR.KAT.
2246	Tülay	A1	4
3245	Semih	A2	4
6226	Yavuz	B1	9
2224	Biröl	A2	4
8292	Burak	A1	4
6242	Yaşar	A2	4
3269	Cem	B2	9
6245	Ali	B3	9

Bileşik anahtara bağlı sahalara da, ayrı bir İNF veri seti oluştururlar (Tablo-6).

Tablo-6: İkinci Normal Form Veri Seti (No.:3)

PROJE KODU	PER.NO.	KATILIM TARİHİ	TAH.SÜR.
VT01	2246	1.3.94	24
VT01	3245	2.4.94	24
VT01	6226	3.4.94	28
VT01	2224	4.4.94	28
VT01	8292	1.5.94	22
BAK04	6242	1.5.94	6
BAK04	3269	1.6.94	6
BAK04	6245	6.7.94	7

d. Üçüncü Normal Form (ÜNF)

İNF'den ÜNF'a geçişte sahalara arasındaki bağımlılıklar araştırılır. Bu adımda çok dikkatli bir analize ihtiyaç vardır. Veri setlerinin ÜNF normalleştirmesinin yapılabilmesi için sahalara mümkün olan tüm kombinasyonlarını teker teker ele alıp olası veri bağımlılıklarının ortaya çıkarılmasında şu sorular sorulur;

"A sahası B sahasına bağımlı mıdır?". Bağımlılığa karar vermeye çalışırken bize yardımcı olacak soru da;

"A'nın verilen bir değerine karşılık, B'nin sadece bir değeri mi elde ediliyor?"

Tablo-5'teki İNF veri seti yukarıdaki sorular çerçevesinde değerlendirildiğinde sahalara bağımlı olduğu görülebilmektedir. Verilen bir grup dereceye karşılık tek bir katsayı elde edilebilirken, verilen katsayının bir değerine karşılık çok sayıda derece değeri elde edilmektedir. Bu durum "Ücret Katsayısı" sahasının "Derece" sahasına bağımlı olduğunu göstermektedir (Şekil-2) /4/.

İNF Veri Seti (No.:1), ÜNF Veri Seti (No.:1) olarak kabul edilir ve anlatılanlar örneğe uygulanırsa İNF Veri Seti (No.:2), İki ÜNF veri setine dönüşecektir (Tablo-7).

Bağımlı saha İNF Veri Seti (No.:2)'den ayrılarak yeni bir ÜNF veri seti oluşturmuştur. Veri setinin anahtarı, bağımlı anahtardır. Örnekte bu "derece"dir. Bağımlı anahtar aynı zamanda ÜNF Veri Seti (No.:2)'de kalmakta ve o veri setinde "yabancı anahtar" görevini yerine getirmektedir. Ayrıca, İNF Veri Seti (No.:3), ÜNF Veri Seti (No.:4) olarak kabul edilir.

Tablo-9: Elde Edilen ÜNF Veri Setleri

PROJE	PERSONEL	PROJE / PERSONEL	DERECE
Proje Kodu * Proje Tipi Tanıtım	Personel No. * İsim Derece	Proje Kodu * Personel No. * Katılım Tarihi Tahsis Süresi	Derece * Ücret Kat- sayısı

Tablo-10: Normalleştirme Sonunda Elde Edilen Veri Setleri

PERSONEL	PROJE	PROJE/PERSONEL	ŞUBE	DERECE
Personel No. * İsim Derece Adres Şube Kodu	Proje Kodu * Proje Tipi Tanıtım	Proje Kodu * Personel No. * Katılım Tarihi Tahsis Süresi Proje Zamanı	Şube Kodu * Şube İsmi Yeri	Derece * Ücret Kat- sayısı

3. SONUÇ VE ÖNERİLER

Mevcut birçok veri analiz tekniklerinin Veri Tabanı oluşturmaya yönelik uygulama çalışmaları son zamanlarda yoğunlaşmaktadır. İlişkisel veri modelini destekleyen Veri Tabanı Yönetim Sistemleri yardımıyla oluşturulmak istenen bir Veri Tabanı tasarımı ve gerçekleştirilmesinde "Üçüncü Normal Form Veri Analiz Tekniği"nin kullanımının önemli bir yer tutacağı değerlendirilmektedir.

K A Y N A K L A R

- /1/ Data C.J. : An introduction to Database Systems, 1990.
- /2/ Sarbanoğlu,H. : Coğrafi Bilgi Sistemi Geliştirme Yöntemi, Harita Dergisi, Sayı: 105, 1990.
- /3/ SSADM : Third Normal Forms Data Analysis, Training Manual, 1977.
- /4/ Yanar,R. : Jeodezik Veri Tabanı Tasarımı.Yüksek Lisans Tezi, YTÜ. İstanbul, 1994.