

İSTATİSTİKSEL VERİ TABANLARI VE BASIN İŞLETMELERİNDE KARAR DESTEK SİSTEMİ OLUŞTURMAK İÇİN KULLANILABİLECEK BİR İSTATİSTİKSEL VERİ MODELİ

Arş. Gör. Cem S. SÜTÇÜ
MARMARA ÜNİVERSİTESİ
İktisadi ve İdari Bilimler Fakültesi

1. Giriş

Bu çalışmanın amacı, Basın Sektöründe çalışıp, belirli konularda karar vermek durumunda olan her düzeydeki yöneticinin ihtiyaç duyabileceği bir Karar Destek Sistemi (KDS-Decision Support System-DSS) oluşturmaktır. Bu amaç doğrultusunda kullanılabilir uygun ve verimli bir İstatistiksel Veri Tabanı Yönetim Sistemi (İVTYS-Statistical Data Base Management System-SDBMS) modeli hazırlamaktır.

2. Karar Destek Sistemleri (KDS)

Karar Destek Sistemlerinin amacı, bilişim teknolojisini yöneticilerin hizmetine daha iyi bir şekilde sunulmasını sağlamaktır. KDS çeşitli alanlardan kavram ve teknikleri bir araya getirerek karar verme durumunda olan kişilere "destek sağlar". Bu alanlar; bilgisayar, matematik, yönetim ve organizasyon, psikoloji ve işletme gibi çok farklı dalları kapsar. KDS, farklı alanlardaki çalışmaların, çeşitli bilgisayar teknikleri kullanılarak karar verme sürecine katkıda bulunmasını temin eder [1].

İstatistiksel Veri Tabanı Yönetim Sistemleri de bu amaca yönelik olarak kullanılabilen bilgisayar tekniklerinden biridir.

3. İstatistiksel Veri Tabanı Yönetim Sistemleri (İVTYS)

Devlet İstatistik Enstitüsü (DİE) gibi kurumların ülke çapında topladığı, nüfus istatistikleri, demografik istatistikler, sağlık, eğitim ve kültür,

adalet, sosyal güvenlik, tarım, madencilik, enerji, imalat sanayii, inşaat, ulaştırma ve haberleşme, turizm, iç ve dış ticaret, fiyat indeksleri, maliye ve milli hesaplar gibi çok çeşitli konularda ve çok büyük miktarlardaki verinin, çeşitli konularda araştırma yapan kişi ve kuruluşlara faydalı olabilmesi için, kullanıcının kolayca ulaşabileceği, istediği veriyi yığın arasından kısa zamanda doğru bir şekilde süzebileceği bir model içinde sunulabilmesi gerekmektedir. Günümüzde kullanılan Veri Tabanı Yönetim Sistemleri -ya da İVTYS ile ilgilenen araştırmacıların ifadesiyle- Ticari (veya İş) Veri Tabanı Yönetim Sistemleri (TVTYS-Commercial (or Business) Data Base Management Systems-CDBMS) modellerinin yapısındaki birtakım eksikliklerden veya veri modeli düzeyi (data model level) ve kavramsal düzeyi (conceptual level) gibi ayrımları yapamamaları nedeniyle eleştirilmektedir [3].

İVTYS'lerin, gerek kullanım alanları, amaçları, hitap ettikleri kullanıcı kitlesi, gerekse bulundukları veri türü, muhafaza edilmiş biçimi, üzerlerinde yapılan sorgulama ve düzenleme işlemleri ve hacimleri bakımından, TVTYS'lerden farklı olduğu, 80'li yılların başından itibaren ortaya atılmaya başlamıştır [4].

4. TVTYS ve İVTYS Arasındaki Farklar ve Alternatif Öneriler

İleri sürülen farklar şu başlıklar altında toplanabilir:

1. İVTYS'de bulunan veri, sadece sayısal veridir. Bunun yanında TVTYS'de sadece sayısal veri değil alfasayısal veri de bulunur (bir işletmenin stoklarının tutulduğu veri tabanı veya PTT'nin abonelerinin kayıtlarının tutulduğu veri tabanı gibi).

2. TVTYS'lerde veri modeli düzeyi detaylı olarak ele alınmamıştır. Bu konu daha detaylı olarak SDM4S modeli konusunda açıklanacaktır.

3. TVTYS'lerde İlişkisel Veri Tabanı Modeli (İLVTM-Relational Database Model-RDBM) kullanılmaktadır. Bu sistemde, değişkenler ve kategoriler arasındaki ilişkiler tespit edilmekte ve bu ilişkiler gözönüne alınarak, toplanan güncel veri satır ve sütunlardan oluşan iki boyutlu, matris şeklinde bir tablo haline getirilmektedir. Bu model, ticari yazılım şirketleri tarafından oldukça kabul görmüş ve günümüzün belli başlı veri tabanı programlarında kullanılmaktadır. Fakat veri sorgulama yeteneklerinin, SQL (Structured Query Language-Yapısal Sorgulama Dili) gibi günümüzün popüler ve oldukça yetenekli sorgulama dillerinin kullanılmasına rağmen yetersiz kal-

dığı ileri sürülmektedir.

Önerilen alternatifler de iki noktada birleşmektedir:

1. Semantik Yaklaşım (Semantic Approach): İstatistiksel kategori ve değişkenlerin arasındaki karmaşık ilişkiyi daha iyi açıklayabilecek bir nesne-ilişki (Entity-Relationship-ER) yapısı oluşturup, bunu konuşma dilinin anlam düzeyine yaklaştırıp ilişkiyel veri tabanı yönetim sistemi sorgulama dilinde iyileştirmeler yapmak.

2. Nesne Yönelimli Veri Tabanı Yönetim Sistemi Yaklaşımı (NYVTYS-Object Oriented Data Base Management System-OODBMS): Bu oldukça radikal yaklaşıma göre, İLVYYS'deki kalıtsal eksiklikler, çok büyük veri tabanlarının yönetimini güçleştirmektedir. Çünkü, bu tip veri tabanları, NASA'nın uydular aracılığı ile elde ettiği iklim, hava koşulları, astrofizik ve okyanus veri sistemlerine ait veri içeren, değişkenleri arasında çok karmaşık ilişkiler kurulması gereken ve günde eklenen veri miktarı 36 Gbit'e (1 Gbit= 10^9 bit) kadar çıkabilen türdedirler [2].

NYVTYS'de değişkenler, kategoriler ve bunların arasındaki ilişkiler birer nesne olarak belirlenmekte ve daha somut hale getirilmeye çalışılmaktadır. Örneğin, Apple Macintosh bilgisayarlarının işletim sistemi de nesne yönelimli bir yaklaşımın ürünüdür. Bir veri dosyası, klasör nesnesi olarak tanımlanmış ve bu klasörün silinebilmesi için de sisteme bir çöp sepeti nesnesi dahil edilmiştir.

Bu üstün özelliklerine rağmen, NYVTYS kullanan veri tabanı programları henüz çok yenidir. Hatta dünyada da kullanımı tam olarak yaygınlaşmamıştır. Ticari olarak yeni yeni kabul görmektedir [6].

5.İVTYS İçin Hangi VTYS Uygun Olabilir?

Bir İVTYS'nin kullanım alanı, kullanıcı profili çok geniştir. Bu tip bir veri tabanında bulunan veri, kullanıcının amacına uygun şekilde süzüldükten sonra, eğer uygun analiz teknikleri ve sorgulama prosedürleri yeterli değilse, yetkin bir istatistik paket programına veya başka bir veri tabanı programına aktarılabilir ve KDS'de kullanılan matematiksel programlama, marjinal analiz, girdi-çıkı analiz, kuyruk teorisi, envanter teorisi, proje yönetimi (PERT/CPM), simülasyon, güvenilirlik ve kalite kontrol, zaman serileri, ile-

riye dönük tahmin (forecasting) ve faktör analizi gibi yöntemlerin üzerlerinde uygulanabilmesine imkan vermemelidir.

Durum bu açıdan ele alındığında, İLVITYS kullanan programların çok büyük bir avantaja sahip olduğu ortaya çıkmaktadır. Halen kullanılmakta olan tüm gelişmiş veri tabanı programları (dBase, FoxPro, Oracle, Sybase, Paradox, Access gibi), elektronik hesap tabloları (Lotus-123, Excel, QPro gibi) ve istatistik paket programları (SPSS, SAS gibi) İLVITYS kullanmakta veya en azından bu tip veri tabanlarıyla alışveriş yapabilecek şekilde düzenlenmektedir. Günümüzde İLVITYS, bir endüstri standardı durumundadır.

Öte yandan, veri miktarı her zaman NASA veri tabanlarında olduğu gibi çok aşırı miktarda olmayabilir. Kavram düzeyinde sorgulama şekillerinin ve değişkenler-kategoriler arası ilişkilerin iyi kurulabildiği bir modelin bir İLVITYS olabileceği ifade edilmektedir.

Bu sebeple. Basın Sektörü ile ilgili Karar Destek Sisteminin kullanacağı İVTYS'nin modelinin ilişkisel bir model olması, tercih edilmesi gereken bir durumdur.

6. İstatistiksel Veri Modelleri (İVM-Statistical Data Models-SDM)

İVT'lerin, istatistiksel amaçlarla toplanan verilerin tablolaştırılması, istatistiksel verinin araştırmacılar arasında paylaşımının sağlanması gibi nedenlerle geliştirilmeye başlanmasına rağmen, ilk yıllarda kullanıcılar, veri hakkında bilgi sahibi olan kişilerle sınırlı kalıyordu ve bilgisayar kaynakları yetersizdi.

Bilgisayar donanımındaki gelişmeler ve bilgisayar kullanımının popüler olmasıyla bu durum değişmeye başladı. İVT'ler daha fazla alanda kullanılmaya başlandı. ve doğal olarak araştırdığı veri tabanındaki verinin boyutlarından habersiz kullanıcıların sayısı arttı. Sonuçta bu kullanıcılar, kullandıkları veri tabanında hangi verinin olduğu ve onlara nasıl ulaşılacağı konularında zorluklarla karşılaştılar. Ardından da istatistiksel verinin yanlış kullanımı ciddi bir problem haline geldi.

Bu durumu düzeltmek için 1980'lerden itibaren pek çok araştırmacı, İVT'lerin mantıksal yapılarını tanımlayan notasyonların formelleştirilmesi üzerinde çalıştılar. Bu tip notasyonlara İstatistiksel Veri Modeli (İVM-Sta-

tistical Data Model-SDM) denilmektedir.

Genel olarak bir veri modeli, (1) veriyi tanımlayan bir notasyon ve (2) veriyi yönetmek için kullanılan işlemler bütünü anlamına gelmektedir.

Bir istatistiksel veri modelinde, veri tabanında tutulacak olan verinin bir tanımı yapılır. Değişkenler, kategoriler belirlenir. Kavramlar ortaya konur. Çalışmanın bir çerçevesi çizilir. Bu çerçeve içine verinin nasıl yerleştirileceği (veri tabanı dosyasının yapısı) belirlenir. Son olarak da gerçek değerler, belirlenen yapıya uygun bir şekilde yerleştirilir.

Bu amaçla kullanılacak genel yapıdaki İVM'lerin ilki 1981 yılında Chan ve Shoshani tarafından ortaya konulan ve oldukça yetkin bir düzeyde olan SUBJECT adlı modeldir. Daha sonra SAM*, GRASS, MEFISTO, CSM, STORM ve SDM4S gibi modeller geliştirilmiştir.

Bu modellerin hemen hemen tamamı kronolojik olarak diğerinden sonra gelen ve öncekilerin eksik taraflarını telafi eden modellerdir [5].

6.1. SDM4S Veri Modeli (Statistical Data Model Based on 4 Schema Concept-4 Şema İVM-4ŞİVM)

Bu model, diğer İVM'lerden ve TVTYS modellerinden farklılık gösterir. Daha kapsamlıdır. Bu modelin bir uygulaması 1990 yılından beri Japonya Ulusal Arazi Ajansı (NLA) tarafından kullanılmaktadır.

Ticari veri tabanları da dahil olmak üzere tüm diğer İVM'lerde model üç aşamadan oluşmaktadır.

1. Kavramsal Şema Aşaması (Conceptual Schema Level)
2. VT Şeması Aşaması (DB Schema Level)
3. Enstantane Veri Şeması Aşaması (Instant Data Schema Level)

Birinci aşamada İVTYS'de olması beklenen verinin bir tanımı yapılır. İVT'nin mantıksal yapısı oluşturulur. Yani, veri tabanında kavramsal olarak olabilecek nesne dünyası (object world of database) tarif edilir. Bunun dışındaki herhangi bir veri nesnesinin veri tabanında yer alması mümkün değildir.

Örneğin, Şekil 1.'deki 8,376 bir özet veridir. İmalat sektöründe çalışan 39 yaş altındaki erkek işçilerin 1980 yılındaki sayısını göstermektedir.

Yaş, cinsiyet birer özetleyici değişken, imalat sanayii, tüm endüstri, servis sektörü ise birer kategori değişkenidir. Bu tablodaki bilgiler de 5 yıl ara ile toplanmaktadır. Bu özet veri Şekil 2.'deki gibi bir çapraz tablo şeklinde de gösterilebilir.

5 Yıllık Endüstri Verileri	Cinsiyet ve yaşa göre işçiler					İşyeri Sayısı		
	Toplam					Toplam	Bireysel	Ortaklı
		E	K	E	K			
1980								
Tüm Endüstri	68942	68942	68942	68942	68942	8932	4753	4179
Üretim	18611	8376	3257	5584	1397	1430	409	1021
Gıda	2329	2329	1231	354	458	291	116	173
Makina	6523	6523	5879	5781	426	326	49	277
Diğer	9762	9762	1443	2458	5896	813	244	569
Hizmetler	37918	37918	9563	12578	2547	6320	3812	2508
Diğer	12410	12410	3910	6897	9872	1182	532	650
1985								
Tüm Endüstri	69037	21859	15672	22738	8768	9468	4848	4620
.
.

Şekil 1. Beş Yılda Bir Yapılan İşçi-İşveren Sayımı Detaylı Sonuçları

	Değişkenler
Kategoriler	Özet Veri

Şekil 2. Çapraz Tablo

İkinci aşama, veri tabanı dosyasının kendisidir. Bu aşamada veri tabanında güncel olarak bulunan veri tarif edilir. Veri tabanının yapısı oluşturulur. Verinin tipi, alan genişliği tespit edilir.

Örneğin, Yaşlar 39 yaş altı ve 40 yaş üstü olarak gruplanır. Cinsiyet erkek ve kadın olarak gruplanır. Yaşlar sayısal değer, cinsiyet kategorik değer alır.

Üçüncü aşamada, VT'de yer alacak her bir özet veri tarif edilir. Örneğin, "1985 imalat sanayiinde çalışan 40 yaş üstü kadın sayısı 8768 dir." gibi. Bu modelin eksik tarafı, bir İVT için kavram şeması aşamasının, İVT nin kapsadığı alanı tarif ederken yetersiz kalmasıdır.

Örneğin, bir firma elindeki tüm işçilere ait Şekil 1. deki gibi bilgileri istatistiksel olmayan bir veri tabanında (Ticari Veri Tabanında) tutuyor olsun. Bu veri tabanında kayıtlı olan tüm işçiler bu işletmenin işçileri olacağı gibi, VT de yer almayan bir işçi o işletmede çalışmıyordur. Zaten bu da beklenen bir durumdur. Çünkü TVT'nin gerçek ilgi alanı sadece kendisi için gerekli olan bilgileri kapsar. Fakat bir İVT için böyle bir durumun sözkousu olmaması gerekir. Örneğin, İVT, TVT'nin kapsamadığı başka işletmelerin işçilerine ait bilgileri de kapsamalıdır. Ancak bu sayede o veri tabanı çok farklı ve çeşitli ihtiyaçları olan kullanıcılara cevap verebilir. Sonuçta, bir TVT'nin veri dünyası İVT için yetersiz kalmaktadır.

Bu yetersizliği ortadan kaldırmak için dördüncü bir aşama geliştirilmiş, daha açık bir ifade ile kavramsal aşama iki kısma ayrılmıştır. Bu dördüncü aşama Veri Modeli Aşaması (Data Model Level) dir.

Bu aşamada İVT'de bulunan tüm veri (kategori değişkenleri -satırlardaki değişkenler- ve özetleyici değişkenlerin -sütünlardaki değişkenler- kapsamları) tarif edilir. Buna ihtiyaç vardır. Çünkü bir işyerinin veri tabanı statik bir veri tabanı olabilir. Ama istatistiksel Veri Tabanının gelişen/dinamik bir veri tabanı olması gerekir. Kullanıcının aradığı bir veriye İVT'de ulaşması mümkün değilse, o veriye en yakın veriye ulaşmasını sağlamak için düzenlemeler yapılır.

Örneğin, Şekil 1.'deki tabloya ek olarak Şekil 3. deki tablonun da İVT'de yer aldığını düşünelim. Bu tablo, ara yıllarda İşçi ve İşveren Anketi sonuçlarının, cinsiyet ve yaş ayırımı yapılmadan toplanmasıyla elde edilmiş olsun.

Kullanıcı, "1982 yılında imalat sanayiinde çalışan 40 yaş üstü kadın sayısı kaçtır?" sorusuna cevap olarak "Böyle bir veri yok" mesajı almamalı. Onun yerine "imalat sanayiinde çalışan 40 yaş üstü kadın sayısı değerleri 5 yılda bir alınmaktadır. Fakat 1980 yılı için imalat sanayiinde çalışan 40 yaş üstü kadın sayısı 1,397'dir" mesajını alabilmelidir.

YILLAR	Endüstrilere Göre İşçi Sayıları			
	Tüm Endüstri	Üretim	Hizmetler	Diğer
1980	68857	18685	37865	12307
1981	70923	19339	38925	12659
1982	72341	19687	39782	12872
.
.

Şekil 3. Ara Yıllarda Yapılan İşçi ve İşveren Anketi Genel Sonuçları

Böyle bir model bir ilişkiel veri tabanı üzerine kurulabilmektedir.

6.2. Basın Sektöründe Bir İVM Denemesi

Basın sektörü ile ilgili olan araştırmacıların gerekli bilgileri bulabilecekleri bir veri tabanı ve bu bilgileri çeşitli yöntemlerle değerlendirebilecekleri bir Karar Destek Sistemi yapısı oluşturulmasında 4 şema istatistiksel veri modeli kullanılmıştır.

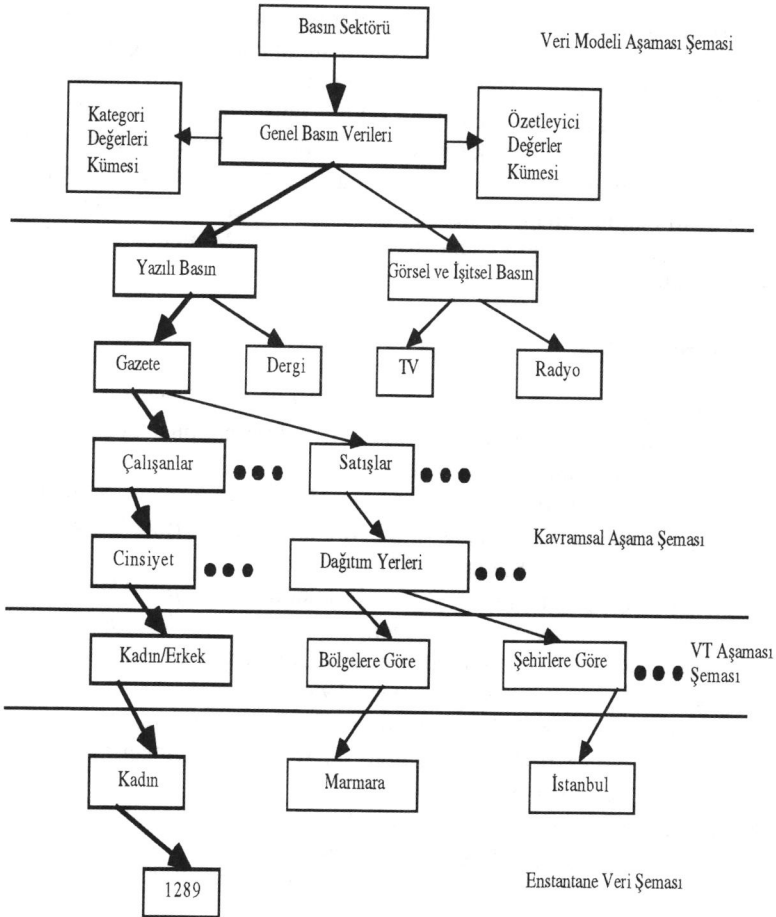
Modelin genel aşamaları şu şekildedir:

- Basın Sektöründe yöneticilerin ihtiyaç duyduğu bilgi kapsamının tespit edilmesi.

- Bu çerçeve içinde, Yazılı Basın, Görsel ve İşitsel Basın gibi üst kategori sınıflarının (superclass) ve çalışanlar, satışlar, çalışanların cinsiyeti gibi alt sınıfların (subclass) kavramsal olarak tespit edilmesi.

- Oluşturulan bu yapıya göre veri tabanına verinin yerleştirilmesi. Örneğin, "Yazılı Basında Gazetelerde çalışan kadın sayısı"nın sorgulanabileceği şekilde veri tabanının yapısının oluşturulması.

- Veri tabanında yer alan tek tek verinin tarif edilmesi. Örneğin, "Yazılı Basında Gazetelerde çalışan kadın sayısı 1289 dur".



Şekil 4. 4 Şema Üzerine Oturtulmuş Basın Sektörü İstatistiksel Veri Modeli Genel Yapısı

Bu aşamalardan ilki olan Veri Modeli Aşamaları, çalışmanın en uzun

sürecek ve en kapsamlı bölümünü oluşturmaktadır. Çünkü bu aşamada basın sektöründe hangi verinin toplanabileceğinin belirlenmesi gerekir. Daha sonra da bu toplanan verinin bir karar verme sürecinde kullanılabilmesi için ne gibi düzenlemelere tabi tutulması gerektiği belirlenecektir.

7. Sonuç

Bir İstatistiksel Veri Modeli kullanılarak Basın Sektöründe yöneticilerin ihtiyaç duyabileceği bir Karar Destek Sistemi hazırlamak için, 4ŞİVM modeli uygun bir seçenek olarak karşımıza çıkmaktadır. Bunun da temel sebebi diğer modellerden daha detaylı bir şekilde oluşturulmuş olmasıdır. İlişkisel veri tabanı yapısını desteklemesi, veri modeli aşamasını da gözönünde bulundurması, kullanıcılar açısından avantajlar taşımaktadır. Çünkü günümüzde ticari olarak kullanılan hemen hemen bütün veri tabanı yönetim sistemleri ilişkisel yapıyı desteklemektedir. Bu da 4ŞİVM modeline esneklik sağlamaktadır.

Bibliografya

- [1] Angehrn, Albert A. *Stimulus Agents: An Alternative Framework for Computer-Aided Decision Making*. 12th International Conference on Decision Support Systems, 1992. s. 81
- [2] French, J. C., *Support for Scientific Database Management*, Statistical and Scientific Databases, (Z. Michalewicz ed.) Ellis Horwood, Chichester, West Sussex, 1991. s. 55.
- [3] Sato, H., Nokano, T., Fukasawa Y., Hotaka R., *Conceptual Schema for a Wide Scope Statistical Database and its Applications*. Proceedings of 3rd. International Working Conference on Statistical and Scientific Database Management, 1986.
- [4] Shoshani A., Olken F., Wong H. K. T., *Statistical and Scientific Database Issues*. IEEE Transactions on Software Engineering, Vol. SD-11, No: 10 1985. ss. 1040-1047.
- [5] *Statistical and Scientific Databases*, Michalewicz Z., ed., Ellis Horwood, Chichester, West Sussex, 1991.
- [6] Stein R. M., *Object Databases*. Byte, Nisan 1994. ss. 75-84.