ANKARA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

KULLANILABİLİRLİK UZMANLARI İÇİN İYİLEŞTİRİLMİŞ WEB UYGULAMASI DEĞERLENDİRME SİSTEMİ

Fatih TEKMEN

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

ANKARA
2019

Her hakkı saklıdır

TEZ ONAYI

Fatih TEKMEN tarafından hazırlanan "**Kullanılabilirlik Uzmanları İçin İyileştirilmiş Web Uygulaması Değerlendirme Sistemi**" adlı tez çalışması 18/06/2019 tarihinde aşağıdaki jüri tarafından oy birliği ile Ankara Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı'nda **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Danışman : Dr. Öğr. Üyesi Ö. Özgür TANRIÖVER
Ankara Üniversitesi / Bilgisayar Mühendisliği Anabilim Dalı

Jüri Üyeleri:

Başkan: Doç. Dr. Çelebi ULUYOL
Gazi Üniversitesi / Bilgisayar ve Öğretim Teknolojileri Eğitimi Anabilim Dalı

Üye : Prof. Dr. Refik SAMET
Ankara Üniversitesi / Bilgisayar Mühendisliği Anabilim Dalı

Üye : Dr.Öğr. Üyesi Ö. Özgür TANRIÖVER
Ankara Üniversitesi / Bilgisayar Mühendisliği Anabilim Dalı

Yukarıdaki sonucu onaylarım.

Prof. Dr. Özlem YILDIRIM
Enstitü Müdür Vekili

ETİK

Ankara Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygun olarak hazırladığım bu tez içindeki bütün bilgilerin doğru ve tam olduğunu, bilgilerin üretilmesi aşamasında bilimsel etiğe uygun davrandığımı, yararlandığım bütün kaynakları atıf yaparak belirttiğimi beyan ederim.

18.06.2019



Fatih TEKMEN

ÖZET

Yüksek Lisans Tezi

KULLANILABİLİRLİK UZMANLARI İÇİN İYİLEŞTİRİLMİŞ WEB UYGULAMASI DEĞERLENDİRME SİSTEMİ

Fatih TEKMEN

Ankara Üniversitesi
Fen Bilimleri Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı

Danışman: Dr. Öğr. Üyesi Ö. Özgür TANRIÖVER

Kullanılabilirlik uzmanlarının ISO 9241-151 standardını referans alan Kamu İnternet Siteleri Rehberi'ne (KAMİS) göre değerlendirme yapabilecekleri bir sistem tasarlanmıştır. İnternet siteleri, uzmanlar tarafından, tasarlanan sistem üzerinden hem Likert tipi ölçeğe göre açılır seçim kutuları ile değerlendirilmiş, hem de bulanık çıkarım yöntemindeki kural kümelerine göre ve daha sezgisel olan sürgü ölçeği ile değerlendirmeye tabi tutulmuştur. Aynı zamanda, uzman kullanıcıların ve normal bilgisayar kullanıcılarının sistem kullanımına dair edindikleri deneyimleri sistem kullanılabilirlik ölçeği (SUS) anketine göre girebilmeleri de sağlanmıştır. İki farklı yöntem ile yapılan uzman değerlendirmeleri ve anketler sonucunda elde edilen veriler kıyaslanmıştır. Toplam üç kamu internet sitesi uzmanlar ve günlük bilgisayar kullanıcıları tarafından değerlendirilmiştir. İlk internet sitesi detaylı değerlendirilmiş ve klasik / sürgü tipi ölçeklerden elde edilen sonuçlar, sistem kullanılabilirlik ölçeği anketinden elde edilen sonuçlar ile kıyaslanmıştır.

Haziran 2019, 76 sayfa

Anahtar Kelimeler: Kullanılabilirlik, KAMİS, değerlendirme, sistem kullanılabilirlik ölçeği anketi, anket

ABSTRACT

Master Thesis

AN IMPROVED WEB APPLICATION EVALUATION SYSTEM FOR USABILITY EXPERTS

Fatih TEKMEN

Ankara University
Graduate School of Natural and Applied Sciences
Department of Computer Engineering

Supervisor: Asst. Prof. Dr. Ö. Özgür TANRIÖVER

A system has been designed for usability specialists to evaluate according to the Public Websites Guidebook (KAMIS), which is based on the ISO 9241-151 standard. The websites are evaluated by the experts according to the Likert scale type selection boxes, and the more intuitive free slide bar with rule sets in the fuzzy inference method. Also, System Usability Scale (SUS) questionnaire was provided for the expert users and non-technical internet users to evaluate their experience in using the websites. The data obtained from the expert evaluations and SUS questionnaire were compared. A total of three public websites were evaluated by experts and non-technical internet users. The first website was evaluated in detail and the results obtained from classical / slider type scales were compared with the results obtained from the system usability scale questionnaire.

June 2019, 76 pages

Keywords: Usability, KAMİS, evaluation, system usability scale questionnaire, questionnaire

ÖNSÖZ ve TEŞEKKÜR

Bu tez çalışmasının gerçekleştirilmesinde bilgi birikimi, fikirleri ve zamanı ile destek olan değerli danışman hocam Ömer Özgür Tanrıöver'e (Ankara Üniversitesi Bilgisayar Mühendisliği Anabilim Dalı) sonsuz teşekkürlerimi sunarım.

Ayrıca uzman ve son kullanıcı olarak tez kapsamında hazırlanan sistemi kullanan ve verilerin üretilmesine katkı sağlayan tüm katılımcılara gönülden teşekkür ediyorum.

Fatih TEKMEN

Ankara, Temmuz 2019

İÇİNDEKİLER

TEZ ONAY SAYFASI
ETİK ..... **i**
ÖZET ..... **ii**
ABSTRACT ..... **iii**
ÖNSÖZ ve TEŞEKKÜR ..... **iv**
KISALTMALAR DİZİNİ ..... **vii**
ŞEKİLLER DİZİNİ ..... **viii**
ÇİZELGELER DİZİNİ ..... **ix**
1. GİRİŞ **1**
1.1 Kullanılabilirlik Nedir? **1**
1.2 Neden Önemlidir? **2**
1.3 Kullanılabilirlik Değerlendirme Yöntemleri **3**
1.4 e-Devlet Sitelerinin Değerlendirilmesi **5**
1.5 Mevcut Durum ve Çalışmanın Amacı **6**
2. LİTERATÜR TARAMASI **8**
2.1 İlke Tabanlı Kullanılabilirlik Değerlendirmeleri **8**
2.2 Likert Tipi ve Sürgü Ölçüm Tekniğine Yönelik Çalışmalar **11**
2.3 Kullanılabilirlik İlkelerin Otomatik Kontrolüne Yönelik Çalışmalar **13**
2.4 Psikolojik Faktörlere Yönelik Araştırmalar **14**
2.5 Genel Değerlendirme **14**
3. MATERYAL **16**
3.1 Kamu İnternet Siteleri Rehberi (KAMİS) **16**
3.1.1 TS EN ISO 9241-151 (İnsan-sistem etkileşiminin ergonomisi) **19**
3.2 Sistem Kullanılabilirlik Ölçeği Anketi (SKÖ) **19**
4. YÖNTEM **21**
4.1 Likert Tipi Ölçek ve Sürgü Ölçeği **23**
4.2 Bulanık Çıkarım Sistemi (BÇS) **25**
4.2.1 Üyelik fonksiyonları **27**
4.2.2 Kurallar **31**
4.3 Araç: Kullanılabilirlik Değerlendirme Sistemi (KDS) **31**
4.3.1 İlke seti yönetimi ekranı **33**
4.3.2 Kullanıcı yönetimi ekranı **34**
4.3.3 Kurum ve internet sitesi yönetimi ekranı **34**
4.3.4 Proje yönetimi ekranı **35**
4.3.5 Proje değerlendirme ekranları **36**
4.3.5.1 Likert tipi klasik değerlendirme ekranı **37**
4.3.5.2 Sürgü ölçeği değerlendirme ekranı **38**
4.3.5.3 Sistem kullanılabilirlik ölçeği anketi ekranı **39**
4.3.6 Rapor ekranları **40**
4.4 Bölüm Özeti **41**
5. BULGULAR ve TARTIŞMA **43**
5.1 Likert Tipi ve Sürgü Ölçeği Değerlendirme Sonuçları **44**
5.2 Likert Tipi Ölçek, Sürgü Ölçeği ve SKÖ Anketi Sonuçlarının Karşılaştırılması **53**
5.3 Uzman ve Son Kullanıcı SKÖ Anketi Değerlendirmeleri **55**

5.4 Tüm Projeler için Sürgü-Bulanık Değerlendirme Sonuçları 56
5.5 Likert Tipi Ölçek ve Sürgü Ölçeği için Korelasyon ve Wilcoxon Testi 60
5.6 Son Puan Analizi 61
5.7 Önem Parametresi Analizi 63
5.8 Uygunluk Parametresi Analizi 65
5.9 Güvenilirlik ve İç Tutarlılık Analizi 67
5.10 Kullanıcı Geribildirimleri 68
5.11 Deney Sonuçlarının Önceki Çalışmalar ile Karşılaştırılması 69
6. SONUÇ 74
KAYNAKLAR 78
EKLER 81
EK 1 SKÖ Anketi Soruları 82
EK 2 Kullanılabilirlik İlkelerinin Otomasyonu 83
ÖZGEÇMİŞ 84

KISALTMALAR DİZİNİ

BİLGEM	Bilişim ve Bilgi Güvenliği İleri Teknolojiler Araştırma Merkezi
ISO	International Organization for Standardization
KAMİS	Kamu İnternet Siteleri Rehberi
KDS	Kullanılabilirlik Değerlendirme Sistemi
SKÖ	Sistem Kullanılabilirlik Ölçeği
TÜBİTAK	Türkiye Bilimsel ve Teknolojik Araştırma Kurumu
WCAG	Web Content Accessibility Guidelines

ŞEKİLLER DİZİNİ

Şekil 4.1 KDS Likert tipi ölçek 24
Şekil 4.2 KDS Sürgü Ölçeği 24
Şekil 4.3 Bulanık Çıkarım Sistemi 25
Şekil 4.4 Girdi Üyelik Fonksiyonu: Önem 28
Şekil 4.5 Girdi Üyelik Fonksiyonu: Uygunluk 29
Şekil 4.6 Çıktı Üyelik Fonksiyonu: Skor 29
Şekil 4.7 KDS aracı ER Diyagramı 32
Şekil 4.8 İlke Seti Yönetimi Ekranı 34
Şekil 4.9 Kurum ve İnternet Sitesi Yönetimi Ekranı 35
Şekil 4.10 Proje Yönetimi Ekranı 36
Şekil 4.11 Klasik Değerlendirme Ekranı 37
Şekil 4.12 Bulanık Değerlendirme Ekranı 39
Şekil 4.13 Sistem Kullanılabilirlik Ölçeği Anketi Ekranı 40
Şekil 5.1 Klasik ve Bulanık Değerlendirme Kategori Ortalamasına göre Radar Grafiği 45
Şekil 5.2 Anasayfa Kategorisine Ait İlkeler için Klasik ve Bulanık Değerlendirme Radar Grafiği 47
Şekil 5.3 Başlıklar Kategorisine Ait İlkeler için Klasik ve Bulanık Değerlendirme Radar Grafiği 48
Şekil 5.4 Yazı Tasarımı Kategorisine Ait İlkeler için Klasik ve Bulanık Değerlendirme Radar Grafiği 49
Şekil 5.5 Arama Kategorisine Ait İlkeler için Klasik ve Bulanık Değerlendirme Radar Grafiği 50
Şekil 5.6 Klasik Uzman Değerlendirmesi Grafiği 51
Şekil 5.7 Bulanık Uzman Değerlendirmesi Grafiği 52
Şekil 5.8 Kategori Bazında Uzman Kullanıcıların Klasik ve Bulanık Değerlendirme Sonuçlarını Gösteren Grafik 52
Şekil 5.9 Klasik / Bulanık Değerlendirme ve SKÖ Anketi Sonuçlarını Gösteren Grafik 54
Şekil 5.10 Uzman ve Son Kullanıcı SKÖ Anketi Sonuçlarını Gösteren Grafik 56
Şekil 5.11 www.ssb.gov.tr internet sitesi ilke bazında bulanık uzman değerlendirmesi grafiği 57
Şekil 5.12 www.uyap.gov.tr internet sitesi ilke bazında bulanık uzman değerlendirmesi grafiği 58

ÇİZELGELER DİZİNİ

Çizelge 5.1 Kategori Ortalamasına Göre Klasik ve Bulanık Değerlendirme Sonuçları . 46
Çizelge 5.2 Ana Sayfa Kategorisi, Klasik ve Bulanık Değerlendirme Sonuçları 47
Çizelge 5.3 Başlıklar Kategorisi, Klasik ve Bulanık Değerlendirme Sonuçları 48
Çizelge 5.4 Yazı Tasarımı Kategorisi, Klasik ve Bulanık Değerlendirme Sonuçları 49
Çizelge 5.5 Arama Kategorisi, Klasik ve Bulanık Değerlendirme Sonuçları 50
Çizelge 5.6 Klasik / Bulanık Değerlendirme ve SKÖ Anketi Sonuçları 54
Çizelge 5.7 Uzman ve Son Kullanıcı SKÖ Anketi Sonuçları 56
Çizelge 5.8 Tüm projeler için bulanık uzman değerlendirmesi sonuçları 58
Çizelge 5.9 Değerlendirilen internet sitelerinin ilkelere göre standart sapma ortalaması 59
Çizelge 5.10 Likert tipi ölçekte ve sürgü ölçeğinde elde edilen ortalama puanlar 60
Çizelge 5.11 Wilcoxon test istatistiği 60
Çizelge 5.12 Likert tipi ölçek ve Sürgü-Bulanık yöntemlerine göre Son Puan analizi .. 62
Çizelge 5.13 Likert tipi ölçek ve Sürgü-Bulanık yöntemlerine göre Önem parametresi analizi 64
Çizelge 5.14 Likert tipi ölçek ve Sürgü-Bulanık yöntemlerine göre Uygunluk parametresi analizi 65
Çizelge 5.15 Likert tipi ölçek ve Sürgü-Bulanık yöntemlerine göre güvenirlik ve iç tutarlılık analizi 67
Çizelge 5.16 Cronbach's α güvenilirlik kategorileri 67
Çizelge 5.17 Çalışmanın Sürgü ölçeği ve Likert tipi ölçeği karşılaştırma çalışmaları ile karşılaştırılması 71

1. GİRİŞ

İnternet sitelerinin veya daha genel bir ifade ile yazılım projelerinin geliştirilmesi aşamasında birçok faktör göz önünde bulundurulmakta, ancak kullanılabilirlik konusu özellikle yazılım geliştirme süreçlerindeki olgunluk seviyesi çok yüksek olmayan özel kurumlar ve kamu kurumları tarafından ihmal edilmektedir. Yazılım sektöründe son yıllarda artan rekabet ortamı, bu alanda faaliyet gösteren firmaların ve kurumların kullanıcı isteklerini; yani kullanılabilirlik konusunu ve kullanıcı deneyimini önemsemelerine, ön plana çıkarmalarına neden olmuştur.

1.1 Kullanılabilirlik Nedir?

Kullanılabilirlik hayatın her alanında geçerli bir kavram olmakla birlikte, bu çalışma kapsamında sadece internet sitelerinin kullanılabilirliği değerlendirilecektir. Bu kavram ile ilgili birçok farklı tanımlama yapılmıştır. Bunların en başta gelenlerinden birisi olarak ISO (International Organization for Standardization) tanımına değinilecektir.

ISO 9241-11:2018 - Ergonomics of human-system interaction -- Part 11: Usability: Definitions and concepts (Anonymous 2018) standardında, kullanılabilirlik başlığı altında sayılabilecek birçok farklı kavram tanımlamalarına yer verilmiştir. Bu standartta, kullanılabilirlik yaygın olarak bilinen "kullanım kolaylığı" ve "kullanıcı dostu olma" ifadelerinden daha kapsamlı bir kavram olarak belirtilmekte ve aşağıdaki tanım yapılmaktadır;

"Kullanılabilirlik, bir sistemin, ürünün veya hizmetin, belirli bir kullanım bağlamında etkinlik, verimlilik ve memnuniyet kavramları ile belirlenen hedeflere ulaşmak için belirli kullanıcılar tarafından ne ölçüde kullanılabildiğidir."

Yine burada yer alan ifadelere göre kullanılabilirlik aşağıdaki konular ile ilişkilidir;

- Kullanıcıların hedeflerine etkili, verimli ve memnun bir şekilde ulaşmalarını sağlamak

- Bir sistemi, ürünü veya hizmeti kullanmaya başladığında yeni kullanıcıların etkili, verimli ve memnun olmalarını sağlamak
- Kullanıcıların her kullanımda sistemle etkili, verimli ve memnun olmalarını sağlamak
- En geniş yetenek yelpazesine sahip kişiler tarafından kullanım
- Risk ve kullanım hatalarının istenmeyen sonuçlarının en aza indirilmesi
- Bakım işlemlerinin etkin, verimli ve memnuniyetle tamamlanmasını sağlamak

1.2 Neden Önemlidir?

Günümüzde internet kullanımı son derece yaygınlaşmış ve internete erişim artık birçok farklı cihazdan, çok farklı bağlamlarda yapılabilir hale gelmiştir. Artık internete erişim eskiden olduğu gibi sadece masaüstü veya dizüstü bilgisayarlar ile değil; cep telefonları, tabletler, televizyon, buzdolabı, fırın, lamba gibi elektrikli ev aletleri, otomobiller, kısacası küçük bir çip takılabilen her tür cihaz tarafından yapılabilir hale gelmiştir. Bu cihazlar kendi aralarında iletişim kurabilir duruma gelmiş; akıllı ev, akıllı kent gibi nesnelerin interneti mimarisini temel alan kavramlar ortaya çıkmıştır.

İnternetin ve interneti temel alan teknolojilerin bu derece hayatın içerisine girmesi, insanların internet sitelerinden kalite beklentisini de yükseltmiştir. İnternet sitelerinin daha fazla insan profiline hitap edebilir olması, kullanım kolaylığı sunması ve insanları düşündürmeyecek, zihnini yormayacak şekilde tasarlanması bir ihtiyaç haline gelmiştir.

Kullanılabilirlik, internet sitelerinin kullanım deneyimini iyileştiren, insanların algılarına ve beklentilerine daha uygun internet sitelerinin üretilmesini sağlayan, nihai olarak kullanım memnuniyetini sağlayan bir unsur olarak, günümüzde çok daha önemli hale gelmiştir. Özellikle son kullanıcı için çok fazla alternatifin olduğu günümüzde, bu konuyu dikkate alan internet siteleri diğerlerinden daha öne çıkmakta ve başarılı olmaktadır.

Mevcut en iyi uygulamalar, bir tasarım projesinde bütçenin %10'unun kullanılabilirlik çalışmalarına tahsis edilmesini önermektedir. Bu bütçe ortalama olarak, bir websitesinin

kalite ölçütlerini iki katından fazla artırır (2,6'lık bir iyileştirme skoru) ve bir intranetin kalite ölçütlerini yaklaşık iki katına artırır (Nielsen 2012).

1.3 Kullanılabilirlik Değerlendirme Yöntemleri

Kullanılabilirlik değerlendirmeleri birçok farklı yöntemle yapılabilmektedir. Kullanılacak olan yöntemin belirlenmesinde değerlendirilmek istenen internet sitesinin niteliği ve genel ihtiyaçlar etkili olacaktır. Farklı koşullarda farklı değerlendirme yöntemlerinin kullanılması çalışmaların daha verimli olmasını sağlayacaktır. Bu yöntemlerden sık kullanılan birkaçına burada değinilecektir.

Göz İzleme Çalışması (Eye Tracking): Bu yönteme göre kullanıcıların bir laboratuvar ortamında, göz izleme cihazının takılı olduğu bir bilgisayar karşısında oturarak, belirli görevleri tamamlamaları istenir. Kullanıcılar bu görevleri tamamlarken göz hareketleri ve tercihe göre ses ve video kayıtları alınır. Kullanıcıların ziyaret ettiği her sayfada göz hareketlerinden hangi noktalara baktıkları, hangi noktalara odaklandıkları ve sayfanın hangi bölgelerinin en fazla dikkatlerini çektiği bilgisi analiz edilir. Burada yapılan analize göre çıkarımda bulunularak test edilen sayfalardaki sorunların tespiti ve düzeltilmesine yönelik kararlar alınır. Bu test yapılırken kullanıcıların kendilerini rahat hissetmelerine dikkat edilmeli ve test öncesinde telkin edilmelidir. (Pernice ve Nielsen 2009)

Yapılan bir araştırmaya göre insanların işlemeyi başardığı bilişsel bilginin %80'inden fazlasının görsel işlemle yapıldığı belirtilmiştir (Wu 2012). Buna göre göz hareketlerinin bilişsel süreçlerde önemli bir bilgi kaynağı olduğu söylenebilir. Göz izleme çalışmaları, göz hareketlerinin kaydedilmesi ve bu verilerin işlenmesi noktasında insan-bilgisayar etkileşimi alanında yapılan çalışmalara katkı sağlamaktadır.

İlk Tıklama Testi: Bu test yöntemi çalışır bir web sayfası veya bir prototip ekran üzerinden gerçekleştirilebilir. Kullanıcıların bir arayüz üzerinde belirlenen bir görevi yapmak için ilk olarak nereye tıkladığı test edilir. (Anonymous 2013)

Sezgisel Değerlendirmeleri ve Uzman Gözden Geçirmeleri: Sezgisel değerlendirmeleri ilk olarak Jacob Nielsen tarafından ortaya atılmış ve günümüzde de yaygın olarak kullanılan yöntemlerden birisi konumundadır. Bu yöntem ilk olarak Molich ve Nielsen (1990) tarafından geliştirilmiş, daha sonra Nielsen (1994) tarafından iyileştirilmiş olan 10 farklı maddeden oluşmaktadır. Bu çalışmanın hızlı ve düşük maliyetle gerçekleştirilebiliyor olması ve etkili sonuçlar üretmesi yaygın kullanımının nedenlerinden olarak görülebilir. Uzman gözden geçirmeleri veya değerlendirmelerinde ise bu sezgisellerin uzmanlar tarafından bilindiği kabul edilir. Bu yüzden resmiyeti daha düşük bir yöntemdir. Uzmanlar potansiyel sorunlar için belirli bir sezgisel setini adreslemek zorunda değildirler (Affairs 2013a).

Nielsen'a ait olan ve en yaygın kullanılan sezgisel seti (Nielsen 1994) aşağıda listelenmiştir;

- Sistem durumunun görünürlüğü
- Sistemin gerçek dünya ile eşleştirilmesi
- Kullanıcı kontrolü ve serbestliği
- Tutarlılık ve standartlar
- Hata önleme
- Hatırlanma yerine tanınma
- Esneklik ve kullanım verimliliği
- Estetik ve minimalist tasarım
- Kullanıcıların hataları tanıma, anlama ve kurtulma işlemlerine yardımcı olma
- Yardım ve dokümantasyon

Sistem Kullanılabilirlik Ölçeği: Bu çalışmada, kullanıcılardan 10 sorudan oluşan bir anketi değerlendirmesi istenir. Kullanıcılar ankette yer alan sorulara "Kesinlikle katılıyorum" ile "Kesinlikle katılmıyorum" arasında değişen 5 farklı cevap verebilirler. Hızlı uygulanabilen, güvenilir bir yöntem olarak sık kullanılanlar arasındadır. İlk olarak 1986 yılında John Brooke tarafından oluşturulmuştur (Brooke 1996). Bu yöntem katılımcıları yönetmek için çok kolay bir ölçeklendirme sunmaktadır. Aynı zamanda

küçük örneklemelerle güvenilir sonuçlar üretmekte ve değerlendirilen internet sitelerinin kullanılabilirliği konusunda etkili bir şekilde ayrım yapılabilmesine imkân tanımaktadır (Brooke 2013).

Bu yöntem aynı zamanda mobil uygulamaların kullanılabilirliği konusunda da tercih edilebilmektedir. 222 genç katılımcının dâhil olduğu bir çalışmada, Android ve iOS sistemlerde en yaygın kullanılan mobil uygulamalar üzerinden SKÖ anketi gerçekleştirilmiş ve değerlendirilen mobil uygulamaların kullanılabilirlik skorları ortaya çıkarılmıştır. Uygulamaların farklı platformlarda olmasının ise önemli bir fark oluşturmadığı görülmüştür (Kaya vd. 2019).

Bu çalışmada, yukarıda sayılan son iki yöntemin kullanılabileceği bir sistem geliştirilerek kullanıcılardan internet sitelerini değerlendirmeleri sağlanacaktır. Daha sonra elde edilen veriler kıyaslanarak bazı çıkarımlarda bulunulacaktır.

1.4 e-Devlet Sitelerinin Değerlendirilmesi

Ticari internet siteleri ile e-Devlet internet siteleri çok genel bir bakış açısı ile değerlendirildiğinde birbirinden çok fazla farklılık arz etmemektedir. Ancak, bu sitelerin kullanılabilirlik açısından değerlendirilmesi, çalışma öncesinde bazı farklılıkların göz önünde bulundurulmasını gerektirir (Wang vd. 2005).

Ticari internet siteleri hedef kitle olarak daha dar bir kesime hitap eder. Bu tür sitelerde kullanıcı farklılıkları daha az olduğu için hedef kitlesine daha odaklanmış ve özelleştirilmiş yapıların kurgulanması e-Devlet sitelerine oranla daha kolaydır. e-Devlet sitelerinde hedef kitle çok daha heterojendir. Kullanıcılar cinsiyet, yaş, eğitim seviyesi, kariyer, gelir gibi açılardan çok büyük varyasyona sahiptir. Bu durum bazı e-Devlet sitelerinde çok fazla bilgi ve hizmet sunulmasına, dolayısıyla kullanıcıların çok fazla algısal ve zihinsel iş yüküne maruz kalmasına neden olabilir.

e-Devlet internet sitelerinin ticari sitelerden farklılaştığı bir diğer nokta ise, bu siteler tarafından sunulan trafik cezası sorgulama, vergi borcu sorgulama gibi hizmetlerin

benzerini başka bir kaynaktan almak mümkün değildir. Bu noktada e-Devlet siteleri, ticari sitelerde olduğu gibi dengi veya rakip bir internet sitesinin baskısı altında değildir. Alternatifinin olmaması, bu internet sitelerini ve hizmetleri sağlayan kamu kurumlarının kullanıcı beklentilerini göz ardı etmesine neden olabilir ve vatandaş açısından dezavantajlı bir durum oluşmasına yol açabilir. Kamu internet sitelerinin vatandaş odaklı hizmetler sunabilmesi için kullanıcı deneyiminin iyileştirilmesine odaklanan kullanılabilirlik çalışmaları yapılmalıdır.

1.5 Mevcut Durum ve Çalışmanın Amacı

Uzman değerlendirmesi türündeki kullanılabilirlik çalışmaları genelde sadece kullanılabilirlik alanında uzmanlığı olan kişiler tarafından yapılabilmekte ve bu kişiler öznel yargılarını değerlendirmelere yansıtmakta, çoğu durumda kullanılan Likert tipi ölçekten dolayı kararlarını özgürce belirtememektedirler. Likert tipi ölçeklerde kişiler belirli ve az sayıda seçenekten bir tanesini seçmek durumunda bırakılmaktadır.

Bu çalışma kapsamında bir araç geliştirilerek kullanılabilirlik konusunda uzmanlığı olmayan fakat web teknolojileri konusunda çalışan, bu alanda teknik uzmanlığa sahip kişilerin kullanılabilirlik değerlendirmesi yapabilmesi sağlanacaktır. Bu yapılan değerlendirmelerde kişilerin bulanık çıkarım sistemi ile desteklenen sürgü ölçeği (bundan sonra "sürgü-bulanık yöntemi" olarak ifade edilecektir) vasıtasıyla gerçek kararlarını daha iyi yansıtabilmeleri ve KAMİS Rehber ilkeleri kapsamında daha objektif bir değerlendirme yapabilmeleri amaçlanmıştır.

Aynı zamanda kişilerin geliştirilen araç üzerinden Sistem Kullanılabilirlik Ölçeği anketini kullanarak internet sitelerini değerlendirebilmeleri de sağlanacaktır. Bu anket, hem teknik uzmanlığa sahip kişiler tarafından, hem de son kullanıcı olarak tabir edilen günlük bilgisayar / internet kullanıcıları tarafından ilgili internet sitesi kullanım deneyimini ölçmek üzere doldurulacaktır. Çalışmada Likert tipi ölçek ve sürgü ölçeği ile yapılan değerlendirmelerden elde edilen ortalama puanların, Sistem Kullanılabilirlik Ölçeği anketinden elde edilen ortalama puana yakınlığı değerlendirilecektir.

Buraya kadar kullanılabilirliğin standartlara göre tanımı, mevcut web teknolojileri üzerinden geliştirilen projeler açısından neden ve nasıl önemli olduğu, günümüzde bu alanda yapılan farklı türde çalışmalar ve e-Devlet sitelerinde bu çalışmaların yapılmasının neden önemli olduğu konularına değinildi. Problem tanımı yapılarak, tez çalışması ile ne tür bir katkı sağlanacağı bu bölümde ele alındı.

2. LİTERATÜR TARAMASI

Geliştirilen ve incelenen sistemin temel aldığı konular üç ana grupta ele alınabilir: Kural veya ilke tabanlı kullanılabilirlik uzman değerlendirme çalışmaları, kullanılabilirlik ilkelerinin kısmen otomatik kontrolüne odaklı çalışmalar ve Likert tipi ölçek ile sürgü ölçeğinin karşılaştırıldığı çalışmalar.

2.1 İlke Tabanlı Kullanılabilirlik Değerlendirmeleri

Kullanılabilirlik değerlendirme yöntemleri üzerine yapılan bir çalışmada (Fernandez vd. 2011), 2703 makaleden oluşan bir ilk set içerisinden toplam 206 araştırma makalesi, haritalama çalışması yapmak üzere belirlenmiştir. Bu makaleler deneysel metotlar ve gözden geçirme metotları olarak iki farklı grupta incelenmiştir. Deneysel metotlar, gerçek son kullanıcıların bir yazılım ürününü belirli görevleri tamamlamak üzere kullanırken, kullanım verilerinin kaydedilmesi ve daha sonra analiz edilmesine dayanmaktadır. Bu tez çalışmasında da kullanılan yöntem olan gözden geçirme metotları ise uzman kullanıcılar tarafından belirli ilke setleri referans alınarak, daha çok kullanıcı arayüzlerinin kullanılabilirlik açısından uygunluğunun gözden geçirilmesine ve değerlendirilmesine dayanmaktadır.

Bir kullanılabilirlik gözden geçirme metodu olan sezgisel değerlendirme yöntemi ilk olarak Jacob Nielsen tarafından bulunmuş ve bu konuda birçok çalışma yapılmıştır. Nielsen ve Molich (1990) tarafından yapılan bir çalışmada, 4 ayrı sezgisel değerlendirme çalışması gerçekleştiriliyor ve bulunan kullanılabilirlik hataları sayısının, 1 ile 5 arasındaki değerlendirici sayısı aralığında çok fazla değiştiği ve 5 kişilik bir değerlendirme ekibinin, kullanılabilirlik hatalarının birçoğunun bulunması için yeterli olacağı sonucuna varılıyor. Bu tez çalışmasında sezgisel olarak nitelenen kurallardan daha net ve öznelliği mümkün olduğunca ortadan kaldırabilmesi amacıyla detaylı teknik yönlendirmelerin bulunduğu KAMİS Rehber referans alınmış ve tüm kullanılabilirlik hatalarının büyük oranda bulunabilmesi için toplam 11 kullanıcı ile değerlendirme çalışması yapılmıştır.

Nielsen (1992) yaptığı çalışmada farklı uzmanlık seviyelerinde ve alanlarında kişilerle sezgisel değerlendirme çalışmaları gerçekleştirmiş ve kullanılabilirlik uzmanlarının bu alanda uzmanlığı olmayan kişilere göre daha iyi sonuçlar ürettiğini görmüştür. Kullanılabilirlik ve arayüz tasarımı konusunda çift uzmanlığı olan kişilerin ise daha da iyi sonuçlar ürettiğine değinmiştir.

Yine Nielsen (1994a, 1994b) kullanılabilirlik değerlendirme yöntemlerine ve sezgisel değerlendirme yöntemine dair yaptığı çalışmalarda, farklı değerlendirme yöntemlerine dair bilgiler vermiş ve etkili bir sezgisel sezgisel seti ortaya çıkarmak için daha önce oluşturulmuş 7 farklı sezgisel seti üzerinden çalışma yaparak nihai iyileştirilmiş bir set belirlemiştir.

Genel olarak sezgisel değerlendirme yönteminin sağladığı bazı avantajlar vardır. Bunlara örnek olarak, hızlı ve daha az maliyetli geribildirim sağlaması, tasarım sürecinin ilk aşamalarında geribildirim alınabilmesi, diğer kullanılabilirlik yöntemleri ile beraber kullanılabilmesi gibi konular sayılabilir.

İnsan – bilgisayar etkileşimi ve kullanılabilirlik konusunda yayınlanan uluslararası standartlar ve geliştirilen bazı ilke setleri üzerine farklı çalışmalar yapılmıştır. Bu çalışmalardan ilkinde kullanılabilirlik üzerine hazırlanan ISO standartları incelenmiş ve karşılaştırmalar yapılmıştır (BEVAN 2001). Daha sonraki yıllarda yapılan bir çalışmada ise üç farklı ilke seti karşılaştırılmıştır (Bevan 2005). Bunlar, Amerika Sağlık ve İnsan Servisleri Departmanı tarafından geliştirilen Araştırma Tabanlı Web Tasarımı ve Kullanılabilirlik İlkeleri seti, bir ISO standardı için geliştirilmiş ilke seti ve İngiltere'deki akademik internet siteleri için geliştirilmiş bir ilke setidir. Bu çalışma sonucu yapılan analizde her üç setin de farklı alanlarda birbirlerine karşı üstünlüklerinin olduğu görüldü. Bir diğer çalışmada (Bevan ve Spinhof 2007) ise ISO 9241-151 standardı ile Amerika Sağlık ve İnsan Servisleri Departmanı tarafından geliştirilen Araştırma Tabanlı Web Tasarımı ve Kullanılabilirlik İlkeleri seti karşılaştırıldı ve bunların kapsamı tartışıldı. Bu çalışmaya göre ISO standardında yer alan ilkelerin %56'sının, diğer ilke setinde yer alan ilkelerin ise %76'sının benzersiz olduğu görüldü.

Rehberler üzerinden kullanılabilirlik değerlendirmesinin yapıldığı iki farklı çalışma bulunmaktadır. Bunlardan ilkinde Türkiye'deki 33 kamu kurumu web sitesi içerik analizi yoluyla değerlendirilmiş, sekiz kamu kurumunun web sitesinden sorumlu kişilerle görüşmeler yapılmıştır (Durmuş ve Çağıltay 2011). Değerlendirme aşamasında yöntem olarak TÜRKSAT tarafından hazırlanan "Kamu Kurumları İnternet Sitesi Önerileri Rehberi – KAKİS" kapsamında oluşturulan ve kamu kurumlarına kendi web sitelerini değerlendirmeleri için verilen araç referans alınmış ve üzerinde bazı uyarlamalar yapılmıştır. Bu araçta her bir işlev sorularla ölçülmüş ve sorulara verilen cevaplar "Evet", "Hayır", "Kısmen" ve "Uygun Değil" olmak üzere dört farklı şekilde seçilmiştir. Ayrıca değerlendirmenin yapıldığı her bir kategori için toplamda %100'e eşit olacak şekilde ağırlıklandırma yapılmıştır. Çalışmada anket/görüşme aracı da kullanılmıştır. Toplam 21 sorudan oluşan bir anket hazırlanmış ve kamu kurumlarının bu anket içerisinde belirlenen farklı konu başlıkları altında görüşleri alınmıştır. Değerlendirme sonucunda web sitelerinin performansları %61,2 ile %88,8 arasında çıkmıştır. Çalışma sonucunda performans puanlarına göre kamu internet siteleri değerlendirilmiş ve önerilerde bulunulmuştur.

Yukarıda bahsedilen çalışmalardan ikincisinde (Arsoy vd. 2013), ISO-9241-151 standardı, Amerika Sağlık Bakanlığı tarafından geliştirilen "Research-Based Web Design & Usability Guidelines" isimli rehber ve KAKİS Rehber referans alınarak kural seti oluşturulmuştur. Toplam 18 kategori altında 139 kural ve kuralları değerlendirmek için 276 değerlendirme noktası belirlenmiştir. Bu değerlendirme noktalarından 146 tanesinin zorunlu, 130 tanesinin ise şarta bağlı olduğuna karar verilmiştir. Daha sonra bu kuralların değerlendirilebileceği bir araç geliştirilmiştir. Araç üzerinden kural seti yönetimi, proje oluşturma, site değerlendirme, proje izleme ve kullanıcı yönetimi gibi işlemlerin yapılabilmesine imkân tanınmıştır. Bu araç ile değerlendirilmek üzere toplam altı farklı internet sitesi belirlenmiştir. Değerlendirmeler 3 farklı değerlendirici tarafından etkileşimli olarak gerçekleştirilmiştir. Değerlendirmeler sonucunda %77,26 ile %56,65 arasında değişen ortalama puanlar elde edilmiştir. Bu çalışmaya dâhil edilen internet siteleri genel olarak değerlendirilmiş ve önerilere yer verilmiştir.

2.2 Likert Tipi ve Sürgü Ölçüm Tekniğine Yönelik Çalışmalar

Ölçüme yönelik geçmişte yapılan çalışmalardan en bilineni Likert ölçüm tekniğidir (Likert 1932). Kağıt-kalem üzerinden ve günümüzde ağırlıklı olarak bilgisayar uygulamaları üzerinden kullanılan bu teknik, deneyim yoğunluğunun doğrusal olduğunu varsayar. Yani tutumların "kesinlikle katılmıyorum" ile "kesinlikle katılıyorum" gibi bir süreç aralığında ölçülebileceğini iddia eder (McLeod 2008). Likert ölçeğinde 5 veya 7 noktalı ölçek kullanılarak bireylerin belirli bir ifadeye ne kadar katıldıklarını veya katılmadıklarını ifade etmelerine imkân tanınır.

Likert tipi ölçek ile sürgü ölçeğinin karşılaştırıldığı bir çalışmada (Cape 2009), farklı ülkelerden bireylerle önce Likert tipi ölçekte 4 soru için değerlendirme yapmaları istendi. Daha sonra bu kişilerin yaptıkları değerlendirmeyi tekrar gözden geçirmeleri ve değiştirmek istedikleri cevapları güncellemeleri istendi. Bu çalışmaya katılan kişilerin birçoğunun verdikleri cevapları değiştirerek güncelledikleri görüldü. Bu çalışmaya göre Likert ölçeğinde bazı sorunların olduğu ifade edildi. Bunlardan ilki, kişi iki farklı soruda örneğin "Kısmen Katılıyorum" cevabını verdiğinde, her iki soruda da aynı seviyede mi katıldığının tam olarak belirlenememesi ve ince farklılıkların bu ölçekte ifade edilememesi olmuştur. İkincisi ise, kişiler örneğin "Kısmen Katılıyorum" ile "Kesinlikle Katılıyorum" gibi cevaplar arasında bir görüşe sahip olduğunda bunu ifade edememesi ve kendisine sunulan cevaplardan birisini işaretlemek zorunda kalmasıdır.

Likert tipi ölçek ile yapılan bu çalışma içerisinde ayrıca farklı görünüm ve özelliklerde sürgü ölçeği ile testler de gerçekleştirildi. Farklı özellikteki sürgü ölçeklerinde, katılımcılardan farklı cevaplar geldiği gözlemlendi. Buna göre sadece değerlendirme yönteminin değil, bu yöntemde kullandığımız aracın görünüm ve özelliklerinin de kişilerin verdiği kararda etkili olduğu ortaya çıkarıldı. Çalışma sonunda niteliksel bir anket çalışması yapılarak katılımcılara geleneksel Likert tipi ölçek ve sürgü ölçeği konusunda çeşitli sorular sorularak geribildirim toplandı. Katılımcılar genel olarak sürgü ölçeğini kullanmaktan daha çok zevk aldıklarını ve bu yöntemi beğendiklerini belirttiler.

Bununla birlikte, sürgü ölçeğinin, görsel analog ölçek ve radyo düğmesi kullanımı ile karşılaştırıldığı bir çalışmada sürgü ölçeği kullanımı ile ilgili olumsuz etkilerden bahsedilmiştir (Funke 2016). Sürgü ölçeğinin, özellikle mobil cihazlarda yanıt oranını, değer dağılımını ve yanıt süresini artırdığı belirtilmiştir. Sürükle-bırak tarzı ölçeklerle, radyo düğmesi gibi üzerine gel-tıkla tarzı ölçeklerin karşılaştırıldığı bu çalışmada genel olarak radyo düğmesi ve görsel analog ölçek kullanımının daha uygun olduğu belirtilmiştir. Benzer özelliklerin karşılaştırıldığı bir başka çalışmada ise masaüstü bilgisayar, tablet ve mobil cihazlarda, 5, 7 veya 11'li ölçeklerde cevap formatlarının bar veya düğme şeklinde kullanıcılara gösterildiği bir deney yapılmıştır (Toepoel ve Funke 2018). Bu çalışmada ortalama skorlar ve cevaplanmayanlar üzerinden farklılıklar incelenmiştir. Masaüstü bilgisayarlarda cevaplanmama oranının daha düşük olduğu, sürgü ölçeğinin, düğme kullanımına göre daha düşük ortalama puan ve daha fazla cevaplanmama oranına sahip olduğu belirtilmiştir. Buradaki sonuçlara göre sürgü ölçeği kullanımının diğer yönteme göre daha zor olduğu çıkarımı yapılmıştır. Ayrıca 11 noktalı ölçeğin 5 noktalı ölçeğe göre daha az cevaplanmama oranına sahip olduğu ve kısa ölçeklere göre daha olumlu sonuçlandığı belirtilmiştir.

Geleneksel radyo buton türü ölçeklerle, sürgü ölçeklerin karşılaştırılmasına yönelik yapılan bir çalışmada (Roster vd. 2015), argüman olarak sürgü ölçeklerin geleneksel yönteme göre daha az tekrarlayan yapıda olması ve kullanıcılar açısından daha eğlenceli bir deneyim sunması temel alınmıştır. Buradaki varsayım, daha etkileşimli bir deneyimin anket yorgunluğunu ve cevaplanmama oranını düşürmesi, bununla birlikte daha kaliteli verinin elde edilmesi olmuştur. Çalışma kapsamında geçmişte yapılan birçok araştırmaya ait deneysel veri incelenmiş ve bir anket gerçekleştirilmiştir. Yapılan ankette geleneksel yöntem ile sürgü ölçeği arasında yukarıda bahsedilen iki argümanı istatistiki olarak destekleyecek yeterli kanıt bulunamamıştır. Cevaplama oranı, tamamlama süresi ve veri kalitesi açısından istatistiki olarak anlamlı bir farklılık olduğunun söylenebilmesi için yeterli veri elde edilemedi. Çalışmada radyo buton formatındaki değerlendirmeden daha yüksek ortalama puan üretildiği görüldü. Bir blog yazısında, Dobronte (2012) buradaki sonucu destekler nitelikte, her iki yöntem arasında belirgin bir farklılığın olmadığını, bunun metottan ziyade tasarımsal bir tercih olduğunu belirtmiştir.

Bu çalışmada yararlanılan bulanık çıkarım yöntemi için üyelik fonksiyonları Likert tipi ölçekte olduğu gibi eşit aralıklar dikkate alınarak oluşturulmuştur. Kural tanımları ise uzman bilgisi ve deneyimine göre tanımlanmıştır. Kullanılabilirlik değerlendirmesinin yapıldığı bir çalışmada bulanık mantık yönteminden faydalanılmıştır (Hub ve Zatloukal 2008). Bu çalışmada araştırmacılar kullanıcıların belirsiz ifadelerle dolu doğal dillerini kullanarak bir yöntem geliştirmeye çalışmışlardır. Bulanık mantık yöntemi belirsizlik söz konusu olduğu için kullanılmıştır. Kullanıcılardan edinilen belirsiz değerlendirme ifadeleri parçalanarak gruplanmış ve belirli değerlendirme ifadelerine dönüştürülmüştür. Bulanık kuralların tanımlanmasında ise uzman deneyimi ve bilgisinden faydalanılmıştır.

2.3 Kullanılabilirlik İlkelerin Otomatik Kontrolüne Yönelik Çalışmalar

Referans alınan farklı ilke setlerine göre kullanılabilirlik değerlendirmelerinin otomatik olarak yapılabilmesine yönelik farklı çalışmalar da gerçekleştirilmiştir. Otomatikleştirilmiş kullanılabilirlik değerlendirmesi, bu çalışma kapsamında da dikkate alınmıştır, ancak bu konu başlı başına ayrı bir çalışma da ele alınabilecek kadar geniş kapsamlı olduğu için mevcut çalışmanın devamı niteliğinde yapılabilecek bir çalışma önerisi olarak sunulmuştur.

Kullanılabilirlik ilkelerinin otomatikleştirilmesi, ilke bazında ele alınması gereken bir süreci gerektirmektedir. Her ilke, tanımlanacak basit kurallarla doğrulanabilecek nitelikte değildir. Bu tür kurallar sadece somut ilkeler için geçerli olup, soyut ifadelerle tanımlanmış ilkeler için daha karışık ve yapay zekâ alanına giren bir çalışma ihtiyacı ortaya çıkmaktadır.

Kullanılabilirlik testlerinin otomatikleştirilmesindeki güncel eğilimlere yönelik yapılan bir araştırmada (Bakaev vd. 2017), otomasyon çalışmaları 3 farklı grup altında toplanmıştır. Bunların ilki, kullanıcıların internet siteleri ile etkileşiminden elde edilen verilerin referans alındığı yöntem olan etkileşim-tabanlı kullanılabilirlik değerlendirmesidir. İkinci gruplandırma ise bir internet sayfasındaki metin öğelerinin yoğunluğu veya metin öğelerin görsel öğelere oranı gibi metriklerin tanımlanmasını ve

ölçülmesini temel alan metrik-tabanlı kullanılabilirlik değerlendirmesidir. Son gruplandırma altında ise yapay zekâ metotları sayesinde gerçek etkileşim olmadan görev modelleri ile kullanım bağlamının (kullanıcı, platform, ortam) kurgulanabildiği model-tabanlı kullanılabilirlik değerlendirme yöntemidir.

2.4 Psikolojik Faktörlere Yönelik Araştırmalar

Duygu durumunun karar verme üzerine etkilerini araştıran bir çalışmada (de Vries vd. 2008), insanların duygu durumu değişiminden sonra kararlarını tekrar gözden geçirmeleri istendi. Bu araştırma, daha önce gerçekleştirilen ve insanların üzgün ruh halinde daha temkinli ve bilinçli değerlendirerek karar verdiklerini, mutlu ruh halinde ise daha sezgisel, hislerine güvenerek ve kısa sürede karar verdiğini ortaya koyan çalışmalar üzerine kurgulandı. Yapılan araştırma sonucunda ruh hali değişen ve önceki kararını tekrar gözden geçiren kişilerin verdikleri kararın öznel değerinde iyileşme gözlemlendiği belirtilmiştir.

Aynı değerin farklı ölçüm birimlerinde rakamsal olarak düşük veya yüksek olarak ifade edilmesinin insanların algısında ve karar verme mekanizmasında farklı etkilerinin olduğuna dair yapılan araştırmalar vardır (Bagchi ve Davis 2016). Örneğin 100g olarak ifade edilen bir ölçü, 100.000mg olarak ifade edilen bir ölçüden daha az olarak algılanmaktadır. Bu araştırma, her iki şekilde ifade edilen miktarın birbirine eşit olmasına rağmen, insanlar tarafından büyük olan rakamsal değerin daha fazla algılandığını ifade etmektedir. Bu tez çalışmasının bulanık değerlendirme bölümünde kullanılan sürgü ölçeğinde, kullanıcılar rakamsal bir değer görmeden, sadece sürgüyü kaydırarak değerlendirme yapmaktadır. Bu araştırma, sürgü ölçeğinin 0 ile 4 veya 0 ile 100 arasında olması kullanıcının vereceği puanı etkileyeceğini göstermektedir. Bundan dolayı çalışma kapsamında sürgü ölçeği gösterilmemekte, ancak arkaplanda sürgü ölçeğinin rakamsal karşılığı 0 ile 100 arasında değişmektedir.

2.5 Genel Değerlendirme

Bu çalışma kapsamında gerçekleştirilen literatür taramasında, sezgisel ve ilke tabanlı değerlendirme, Likert tipi değerlendirme, sürgü ölçeğine göre değerlendirme ve

kullanılabilirlik ilkelerinin otomatik değerlendirilmesine yönelik daha önce yapılan çalışmalar konusunda araştırma gerçekleştirilmiştir. Literatür taraması bölümünde, bu çalışmaların gruplanarak alt başlıklar halinde daha düzenli ve anlaşılır şekilde sunulması amaçlanmıştır.

Tez çalışmasında da ele alınan yöntem olarak, kullanılabilirlik değerlendirmelerinin hızlı ve maliyet açısından gerçek sistem kullanıcılarıyla yapılan çalışmalara oranla çok daha düşük olabilmesi amacıyla ilke değerlendirmeleri üzerine yoğun araştırma gerçekleştirilmiştir. Bu tür değerlendirmelerin yapılabileceği farklı yöntemler olan Likert tipi ölçeğe ve sürgü ölçeğine göre daha önce yapılan çalışmalar incelenmiş, her iki yöntemin zayıf ve güçlü yanları ele alınmıştır. Aynı zamanda ilkeler üzerinden otomatik değerlendirme yapılabilmesine yönelik daha önce yapılan çalışmalar incelenmiş ve değerlendirme esnasında kullanıcı kararlarını etkileyebilecek psikolojik faktörlere yönelik araştırma çalışması gerçekleştirilmiştir.

Bu çalışma kapsamında ele alınan ilkelere, standartlara göre değerlendirme yöntemi bütün internet sitelerinde kullanılabilir özellikte olup, bu yöntemin tercih edilmesinde genellikle hızlı çözüm üretmesi ve maliyetinin ucuz olması etkili olmaktadır. Giriş bölümünde çalışmanın amacına dair mevcut çalışmaların eksiklikleri, problemleri ortaya konmuş ve bu tez çalışması kapsamında ne tür iyileştirme ve çözüm geliştirileceği anlatılmıştır.

3. MATERYAL

Kullanılabilirlik değerlendirmelerinde öznel veya kişisel bakış açısının sonuç üzerindeki etkisi fazla olmaktadır. Bu konuda her ne kadar yayımlanmış standartlar, rehberler ve konu uzmanı kişilerin tanımladığı ilkeler, kurallar olsa da değerlendiren kişinin deneyimi, önyargıları ve bakış açısı gibi öznel parametreler, sonuç üzerinde ciddi farklılıklar oluşturabilmektedir.

Bu çalışmada, öznel parametrelerin kişilerin değerlendirme sonucu üzerindeki etkisini daha aza indirgeyebilmek için onlara yardımcı olacak ve karar verebilmeleri noktasında referans olarak kullanabilecekleri nitelikte bir rehber olarak, Kamu İnternet Siteleri Rehberi (KAMİS) kullanılmıştır. Bu rehberi diğerlerinden farklı kılan nokta, her ilke için uzmanlar tarafından yazılan yönergeler ile kullanıcıların daha somut değerlendirme yapabilmesine imkân tanınmasıdır.

3.1 Kamu İnternet Siteleri Rehberi (KAMİS)

KAMİS Rehber, kamu internet siteleri için TÜBİTAK BİLGEM Yazılım Teknolojileri ve Araştırma Enstitüsü tarafından hazırlanmış bir rehberdir. Bu rehberde kullanılabilirlik ve erişilebilirlik konularında kamu internet sitelerinde yer alan temel sorunlar hedeflenmiş ve bunların düzeltilmesine katkı sağlayacak nitelikte yönlendirici ilkelere ve yönergelere yer verilmiştir. Kullanılabilirlik ilkelerinin hazırlanması aşamasında TS EN ISO 9241-151 (İnsan-Sistem Etkileşiminin Ergonomisi) standardı (Anonim 2008) referans alınmış, erişilebilirlik bölümünde ise ISO/IEC 40500:2012 (Web İçeriği Kullanılabilirlik Standartları ve Kriteri) standardı (Anonymous 2012) referans alınmıştır.

KAMİS Rehber ilk olarak 2014-2016 yılları arasında oluşturulmuştur. Projenin bu aşamasında TÜBİTAK'ta çalışan konu uzmanları ile akademik alanda kullanılabilirlik ve erişilebilirlik konusunda ihtisas yapmış olan akademisyenler yer almıştır. Proje sonunda 6 bölümden oluşan bir kitap ortaya çıkmış ve bu kitap kamu kurumlarına dağıtılarak kullanılabilirlik ve erişilebilirlik konusunda farkındalık oluşturulmaya çalışılmıştır. Proje çıktısı aynı zamanda www.kamis.gov.tr internet sitesinden PDF

formatında çevrimiçi herkesin erişimine açılmıştır. Rehberin ilk sürümünde aşağıda verilen bölümler yer almıştır;

1. İnternet Sitelerinde Kullanılabilirlik
2. Genel Özellikler
3. Görsel Tasarım ve Kullanıcı Arayüzü Özellikleri
4. Erişilebilirlik
5. Kamu İnternet Siteleri Mevzuat Bilgileri
6. Kullanılabilirlik Testleri ve Uygulamaları

Rehberin sonunda; TS EN ISO 9241-151 Eşleştirme Listesi, Rehber İlkeleri ve TS EN ISO 9241-151 Eşleştirme Listesi, TS EN ISO 9241-151 Kontrol Listesi Şablonu ve WCAG 2.0 Eşleştirme Listesi ek olarak verilmiştir.

Rehberin bu sürümünde daha çok kitabi bilgilere yer verilmiştir. Genel hatları ile nelerin yapılabileceği anlatılmış ancak çok fazla teknik detaya girilmemiştir. Rehberin geliştirilmesi için projenin ikinci fazı başlatılmış ve kamu kurumları ve özel sektörden firmalarla odak grup çalışmaları gerçekleştirilmiştir. Yapılan bu çalışmalar sonucunda, kamu kurumu çalışanlarının KAMİS Rehber'de daha uygulamaya dönük ve pratik bilgilerin yer almasına yönelik taleplerinin olduğu anlaşılmıştır. İkinci faz çalışması sonucunda özellikle kullanılabilirlik bölümünde her ilke için ayrıca çalışma yapılarak bu ilkelerin daha uygulanabilir hale dönüştürülebilmesi amacıyla yönergeler eklenmiştir.

Projenin ikinci fazında, yine akademisyenlerden bu alanda ihtisas yapan kişilerin de katkıları ile rehber daha detaylı ve ihtiyacı karşılayacak şekilde güncellenmiş olup toplam 5 bölüm olarak tasarlanmıştır. Rehberde yer alan bölümler aşağıdaki gibidir;

1. İnternet Sitelerinde Kullanılabilirlik ve Erişilebilirlik
2. Kullanılabilirlik İlkeleri
3. Erişilebilirlik İlkeleri
4. Kamu İnternet Siteleri Mevzuat Bilgileri

5. Teknoloji Rehberi

Kullanılabilirlik bölümünde toplam 18 farklı kategori altında ilkeler ve bu ilkeleri açıklayan yönergelere yer verilmiştir. Faz 2 kapsamında gerçekleştirilen en büyük değişiklik, her ilke için detaylı yönergelerin yazılması olmuştur. Rehberde yer alan tüm ilke kategorilerinin listesi aşağıda görülebilir;

1. Ana Sayfa
2. Arama
3. Aygıt Bağımsız Tasarım
4. Bağlantılar
5. Başlıklar
6. Bilgi Güvenliği
7. İçerik Politikasının Belirlenmesi
8. Kullanıcı Odaklı Tasarım
9. Kurumsal İletişim
10. Mobil Odaklı İlkeler
11. Multimedya Elamanları
12. Navigasyon
13. Performans Odaklı Tasarım
14. Sayfa Yapısı
15. Sitenin İsimlendirilmesi
16. Teknik ve Altyapısal Etkenler
17. Veri Girişi
18. Yazı Tasarımı

Kategoriler önemine ve kapsamına göre farklı sayıda ilkeye ve detaylı yönergelere sahiptir. Aynı zamanda bu rehberde yer alan ilkeler, akademisyenler ve konu uzmanı kişiler tarafından önem derecesine göre 5 yıldız üzerinden puanlanmıştır. İlkelerin tamamı önemli olarak değerlendirilmiş ve verilen puanlar, ilkeler arasında orantısal olarak ele alınmıştır.

Bu çalışmada, rehberin kullanılabilirlik ilkelerinin bulunduğu bölümü referans alınmıştır. Yönergelerin eklenmesi ile yönlendirici ve detaylı ilkeler elde edilmiştir. Ancak bu çalışmada, uzman kullanıcıların sıkılmadan, odaklı bir değerlendirme yapabilmesi için seçilen kategori sayısı 4 ile sınırlı tutulmuştur. Bu kategoriler seçilirken değerlendirilecek internet sitelerinin niteliği göz önünde bulundurulmuş ve buna göre uygun olanlar seçilmiştir. Çalışmada kullanılan 4 kategori son bölümde ek olarak verilmiştir.

3.1.1 TS EN ISO 9241-151 (İnsan-sistem etkileşiminin ergonomisi)

Bu standart, web kullanıcı arayüzü tasarımı konusunda üst düzey tasarım kararları ve tasarım stratejisi, içerik tasarımı, gezinme ve arama, içerik sunumu konularına odaklanmaktadır (Anonim 2008). KAMİS Rehber'in kullanılabilirlik ilkeleri bölümünde referans alınan bu standart, internet sitelerinde yapılan değerlendirmenin de genel kapsamını ortaya koymaktadır. KAMİS Rehber içerisinde yer alan kategoriler / gruplandırmalar standart ile birebir uyumlu değildir. Ancak ilke seviyesinde, standart içerisinde yer alan başlıklar ve konulara referans verilmiştir.

3.2 Sistem Kullanılabilirlik Ölçeği Anketi (SKÖ)

Sistem kullanılabilirlik ölçeği anketi bir sistemin kullanılabilirliğini hızlı ama çok kesin olmayan bir şekilde ölçmek için kullanılabilecek güvenilir bir yöntemdir. Anket, toplam 10 sorudan oluşmaktadır. Sorulara verilen cevaplar "Kesinlikle katılıyorum" ile "Kesinlikle katılmıyorum" arasında değişen 5 farklı cevaptan oluşabilir. Aslen 1986 yılında John Brooke tarafından oluşturulan yöntem (Brooke 1996), yazılım, donanım, mobil cihaz, internet sitesi ve uygulamalar dâhil olmak üzere birçok farklı ürün ve servisin, kullanıcıları tarafından değerlendirilmesine yardımcı olmaktadır (Affairs 2013b).

Kullanıcı değerlendirmesinin çok basit ve kullanıcılara kolaylıkla anlatılabilir olması, küçük örneklem sayılarında güvenilir sonuç üretmesi ve internet sitelerinin kullanılabilirliği konusunda hızlı bir şekilde ayrım yapabilmesi, SKÖ anketi kullanımının sağladığı faydalardandır.

Ankette yer alan sorular bir olumlu, bir olumsuz olacak şekilde sıralanmıştır. Bunun nedeni kullanıcıların dikkatini üst seviyede tutmak içindir. Soruların tamamı olumlu veya olumsuz olacak şekilde sıralandığında, kullanıcılar puanlama yaparken daha dikkatsiz davranabilmektedir. Bu şekilde soruyu anlamak için dikkatlerini daha fazla yoğunlaştırma ihtiyacı duymaktadırlar (Brooke 2013).

Anketteki soruların nihai skorları 0 ile 4 arasında değişmektedir. Skorun hesaplanması için olumlu sorulara verilen puandan 1 çıkarılır, olumsuz sorulara verilen puan ise 5'ten çıkarılır. 10 sorudan oluşan anketin sonucu 0 ile 40 puan arasında değişir. Bu puanın 0 ile 100 arası bir puana dönüştürülmesi için 2,5 ile çarpılır.

4. YÖNTEM

Bu çalışmada bir grup teknik uzmanlığa sahip kişiler üzerinden deneysel araştırma yöntemi uygulanmıştır. Kişilerden kullanılabilirlik değerlendirmesini Likert ve sürgü tipi ölçekler olmak üzere iki farklı şekilde gerçekleştirmeleri istenmiş ve değerlendirme sonuçlarından elde edilen ortalama puanlar ile SKÖ anketinden elde edilen ortalama puan karşılaştırılmıştır.

KAMİS Rehber içerisinde yer alan ilkelerin değerlendirilmesi aşamasında kişilerin sürgü tipi ölçek üzerinden verdiği puanlar bulanık çıkarım yöntemi ile kural setlerine tabi tutularak daha somut hale getirilmeye çalışılmıştır.

KAMİS Rehber ilkelerine göre uzman kullanıcıların değerlendirme yaptıkları bölümde, toplam 11 kullanıcıdan belirlenen internet sitelerini değerlendirmeleri istendi. 5 kullanıcı ile yapılan kullanılabilirlik çalışmalarının mevcut kullanılabilirlik sorunlarının %85'ini bulduğuna dair araştırmalar yapılmıştır (Nielsen 2000). 15 kullanıcı ile yapılan çalışmalarda ise tüm kullanılabilirlik sorunları tespit edilebiliyor. Bu çalışma araştırmaya göre ideal bir sayı olan 11 kullanıcı ile gerçekleştirilmiştir. Son kullanıcılara yönelik kullanım deneyiminin ölçülmesini sağlayan SKÖ anketi bölümünde ise uzman kullanıcılar da dâhil olmak üzere toplam 30 kullanıcıdan www.turkiye.gov.tr internet sitesini değerlendirmeleri istendi. Değerlendirmeler için sistem üzerinden süre tanımlanabilmekle birlikte, bu çalışma kapsamında süre zorunlu tutulmadı.

Kullanıcılardan aşağıda yer alan 3 farklı siteyi değerlendirmeleri istendi;

1. www.turkiye.gov.tr
2. www.ssb.gov.tr
3. www.uyap.gov.tr

www.turkiye.gov.tr internet sitesi herkes tarafından kullanılan ve kamu internet siteleri arasında hitap ettiği kitle açısından en önemli sayılabilecek internet sitesi olduğu için değerlendirilmek üzere seçildi. www.ssb.gov.tr internet sitesi modern bir tasarıma,

www.uyap.gov.tr sitesi ise güncel tasarım trendlerinden daha uzak bir tasarıma sahip olduğu için değerlendirilmek üzere seçildi. Böylelikle farklı özellikteki internet sitelerinin değerlendirmesinden elde edilen puanların dağılımı ve standart sapması üzerinden önerilen yöntemin tutarlılığı ölçüldü. Seçilen internet sitelerinin tamamının kamu kapsamında olması ise KAMİS Rehber'in kamu kurumlarına yönelik olarak geliştirilmesinden dolayı tercih edildi.

KDS aracı üzerinden her 3 site için oluşturulan projelere, KAMİS ilke setinden 4 farklı kategori altında toplam 28 ilke değerlendirilmek üzere seçildi. Bu kategorilerin seçiminde hem birçok siteye kolayca uygulanabilir olmaları, hem de uzman kullanıcı olarak tanımlanan kişilerin kolayca anlayabilecekleri ilkeleri içermesi açısından dikkat edildi. Her 3 site aşağıda yer alan kategoriler açısından değerlendirildi;

1. Ana Sayfa
2. Başlıklar
3. Yazı Tasarımı
4. Arama

Seçilen internet sitelerini, yukarıda belirtilen 4 kategori altında değerlendirecek olan uzman kullanıcılar uygun örneklem olarak belirlendi. Çalışmanın hedef kitlesindeki kişileri temsil edecek nitelikte, web teknolojileri konusunda uzmanlığı olan ve 11 kişiden oluşan ölçüt bir grup belirlendi. Yukarıda belirtilen 11 kişi de dâhil olmak üzere, SKÖ anketini değerlendiren toplam 30 kişi, uygun örneklem olarak çalışmanın hedef kitlesini temsil edecek kişilerden seçildi. Örneklem sayısının 11 olmasında, her bir kullanıcı ile yapılacak çalışmanın uzun zaman alması ve belirli teknik yetkinlikteki insanlar ile bu çalışmanın yapılabilmesi etkili olmuştur.

Uzman kullanıcı niteliğindeki kişilerin tamamı erkek ve 5 yıldan fazla iş tecrübesine sahiptir. 8 kişi web teknolojileri konusunda yazılım geliştiren kişilerdir. Diğer 3 kişi ise web teknolojileri üzerine farklı konularda çalışmalar yapmaktadır. 11 kişinin yaş ortalaması 33,27'dir.

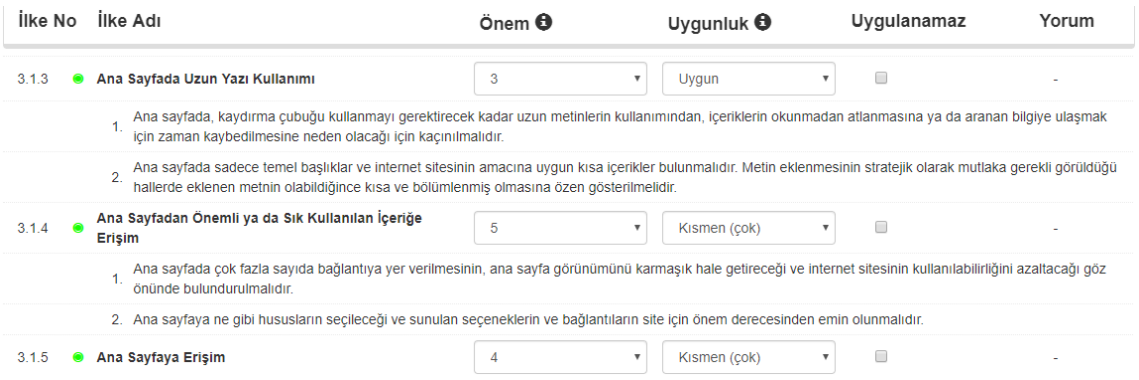
Uzman kullanıcılar ilk proje altında tanımlı olan siteyi hem açılır seçim kutuları vasıtasıyla klasik yöntemle, hem de sürgü-bulanık yöntemi vasıtasıyla değerlendirdiler. Her iki yöntemle yapılan değerlendirme sonuçları karşılaştırıldı. Aynı zamanda, toplam 30 kullanıcının bu site için yaptıkları SKÖ anketi değerlendirmeleri incelenerek, siteyi ilke setine göre de değerlendirmiş olan uzman kullanıcıların ve sadece site kullanım deneyimi olan son kullanıcıların değerlendirmeleri arasındaki farklar araştırıldı.

Diğer iki proje altında tanımlanan internet siteleri sadece uzman kullanıcılar tarafından ve bulanık değerlendirme yöntemi kullanılarak değerlendirildi. Bu iki internet sitesinin ortaya çıkardığı sonucun, ilk projedeki internet sitesinden elde edilen verilerin güvenilirliğine katkı sağlaması amaçlandı.

Yapılan çalışmada uzman kullanıcılar tarafından gerçekleştirilen ve bulanık hesaplama yönteminin dikkate alındığı sezgisel sürgü ölçeği ile yapılan değerlendirmenin, hem klasik yönteme göre daha tutarlı olması ve hem de bu yöntemle, son kullanıcılara yönelik tasarlanmış olan SKÖ anketi sonuçlarına daha yakınsanmış bir ortalama puan elde edilmesi hedeflendi.

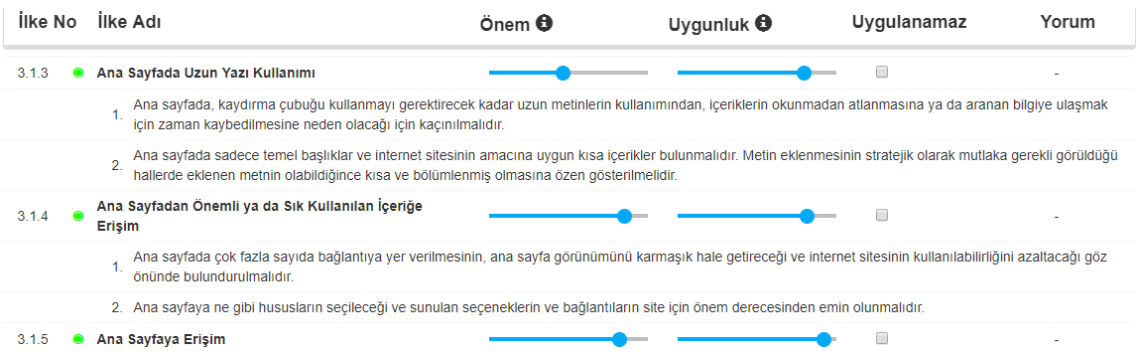
4.1 Likert Tipi Ölçek ve Sürgü Ölçeği

Kullanıcı tutumlarına yönelik yapılan ölçümlerden, günümüzde de aktif bir şekilde kullanılan yöntemlerin en eskilerinden birisi Likert ölçeğidir (Likert 1932). Bu tür ölçeklerde, insanların tutumları ve görüşleri sabit seçim yanıt formatları ile ölçülmektedir. Bu sayede, kullanılan sıralı ölçeklerle, insanların sorulan sorulara ne derecede katıldığı ölçülebilmektedir.



Şekil 4.1 KDS Likert tipi ölçek

Likert tipi ölçek ile sürgü ölçeklerinin karşılaştırılmasına dair birçok çalışma yapılmıştır. Bunlardan bir kısmı Likert ölçeğini savunurken (Funke 2016), bir kısmı da sürgü ölçeğinin bazı noktalarda daha kullanışlı olduğunu ve kullanıcıların daha cazip bulduğunu (Cape 2009) ortaya koyan çalışmalar yapmıştır. Likert ölçeği, aynı değerlendirmeyi yapan kullanıcıları tamamen aynı fikirdelermiş gibi göstermesi ve ara değerlerin çok az olmasından dolayı kişileri belirli seçenekler yapmaya zorlaması gibi dezavantajlara sahiptir. Bu noktalarda sürgü ölçeği kullanımı daha avantajlı olmaktadır.



Şekil 4.2 KDS Sürgü Ölçeği

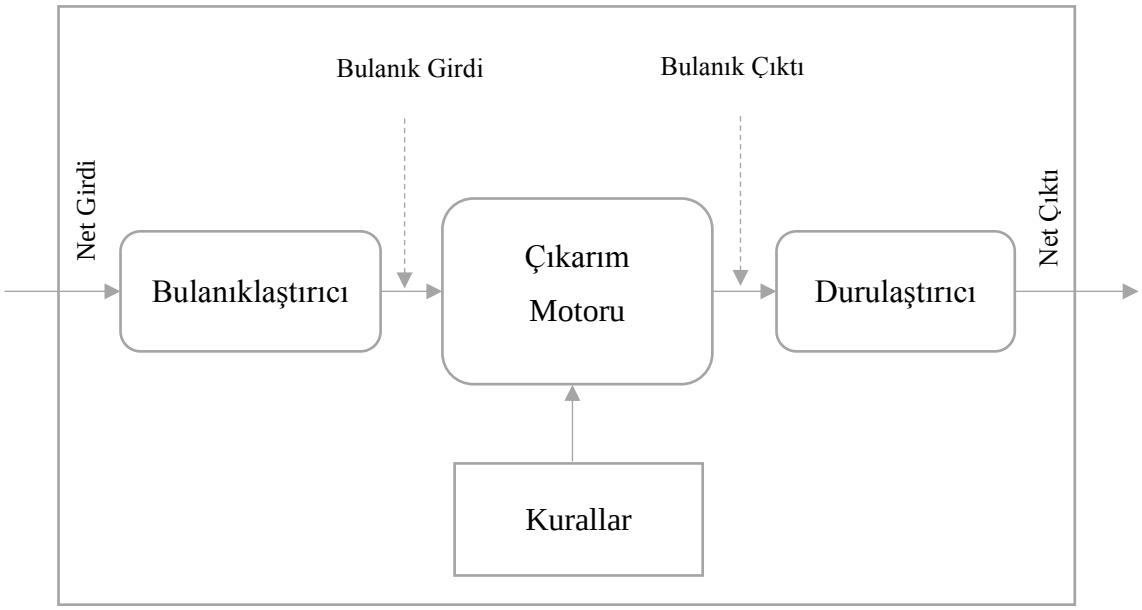
Burada yapılan çalışmada, her iki ölçüm tekniği kullanılarak veri üretilmiş ve bu veriler üzerinden karşılaştırma yapılmıştır. Bulanık çıkarım yöntemi ile desteklenen sürgü ölçüm yönteminin (sürgü-bulanık yöntemi) SKÖ anketinden elde edilen sonuçlara daha yakın değerler ürettiği görülmüştür. SKÖ anketinde yapılan değerlendirmeler; uzmanlar

tarafından yapılan değerlendirmeler ve son kullanıcı değerlendirmeleri olarak sınıflandırılmış ve tutarlılığın gösterilmesi adına sadece uzman kullanıcıların sürgü-bulanık yönteminde ve SKÖ anketinde yaptıkları değerlendirmelerden elde edilen sonuçlar dikkate alınmıştır. SKÖ anketinde son kullanıcılar genel olarak uzman kullanıcılardan daha yüksek puanlar vermişlerdir. Ancak uzman kullanıcıların her iki değerlendirme de birbirine çok yakın puanlar ürettiği görülmüştür.

4.2 Bulanık Çıkarım Sistemi (BÇS)

Bulanık kümeler ilk olarak Lotfi A. Zadeh (1965) tarafından klasik kümelerin bir uzantısı olarak tanıtılmıştır. Bulanık çıkarım yöntemi genel olarak tarif edilmek istenirse, girdi(ler)in bulanık mantık işlemleri kullanılarak çıktıya dönüştürülmesi süreci olduğu söylenebilir. Üyelik fonksiyonları, bulanık mantık operatörleri ve kurallar bu süreçte kullanılan kavramlardandır. Bu çalışmada Mamdani bulanık çıkarım yönteminden faydalanılmıştır. Mamdani yöntemi, insan girdisinin kullanıldığı sistemler açısından daha uygundur. Aynı zamanda sezgisel olmakla birlikte, yaygın kullanıma sahip ve kabul görmüş bir standart halini almıştır.

Basit bir bulanık çıkarım sisteminin yapısı aşağıdaki gibi gösterilebilir;



Şekil 4.3 Bulanık Çıkarım Sistemi

Bulanıklaştırma işleminde net girdiler üyelik fonksiyonlarına göre 0 ile 1 arasında değişen bulanık değerler ile eşleştirilir. Üyelik fonksiyonları, "çok düşük", "düşük", "normal", "yüksek" ve "çok yüksek" gibi bulanık kavramları temsil eder. Her net girdiye, bulunduğu değer aralığında tanımlanmış üyelik fonksiyonuna olan üyelik derecesine göre 0 ile 1 arasında değişen bir değer atanır. Kuralların ilk bölümünde (sol tarafta) yer alan "ve", "veya" gibi bulanık birleştirme operatörleri ile bağlanmış üyelik kümeleri öncül olarak isimlendirilir. Öncül bölümünde birleştirme operatörü olarak "ve" kullanıldıysa buradaki değerler arasından en küçük olanı alınarak kuralın ağırlığı (etkisi) hesaplanır;

$$\mu_{A\cap B} = T\big(\mu_A(x), \mu_B(x)\big)$$

$$\min(\mu_A(x), \mu_B(x))$$

μ_A = A kümesinin üyelik oranı

μ_B = B kümesinin üyelik oranı

Öncül bölümünde "veya" birleştirme operatörü kullanıldıysa değerler arasından en büyük olan alınarak kuralın ağırlığı hesaplanır;

$$\mu_{A\cup B} = T\big(\mu_A(x), \mu_B(x)\big)$$

$$\max\big(\mu_A(x), \mu_B(x)\big)$$

μ_A = A kümesinin üyelik oranı

μ_B = B kümesinin üyelik oranı

Kuralların ikinci bölümünde (sağ tarafta) yer alan çıktı üyelik kümesine ait fonksiyon, sol tarafta hesaplanan kural ağırlık değerine göre kırpılır. Daha sonra tüm kurallardan elde edilen kırpılmış üyelik kümesi fonksiyonları "veya" operatörü kullanılarak birleştirilir ve çıktı dağılımı kümesi elde edilir.

Durulaştırma işlemi için girdi, bir önceki adımda elde edilen çıktı dağılımı kümesi olur. Bu küme üzerinden durulaştırma işlemi yapılarak tek bir net çıktı değeri hesaplanır. Durulaştırma işleminde farklı metotlar uygulanabilir. Ancak bu çalışmada ağırlık merkezi (centroid) metodu kullanılmıştır. Girdi olarak kullanılan bulanık kümenin temsil ettiği alanın ağırlık merkezi hesaplanarak durulaştırma işlemi gerçekleştirilmiş olur.

Ağırlık merkezine göre durulaştırma işlemi aşağıdaki formüle göre gerçekleştirilir;

$$z = \frac{\sum_{j=1}^{q} Z_j u_C(Z_j)}{\sum_{j=1}^{q} u_C(Z_j)}$$

z = Ağırlık merkezi

u_C = C kümesindeki üyelik değeri

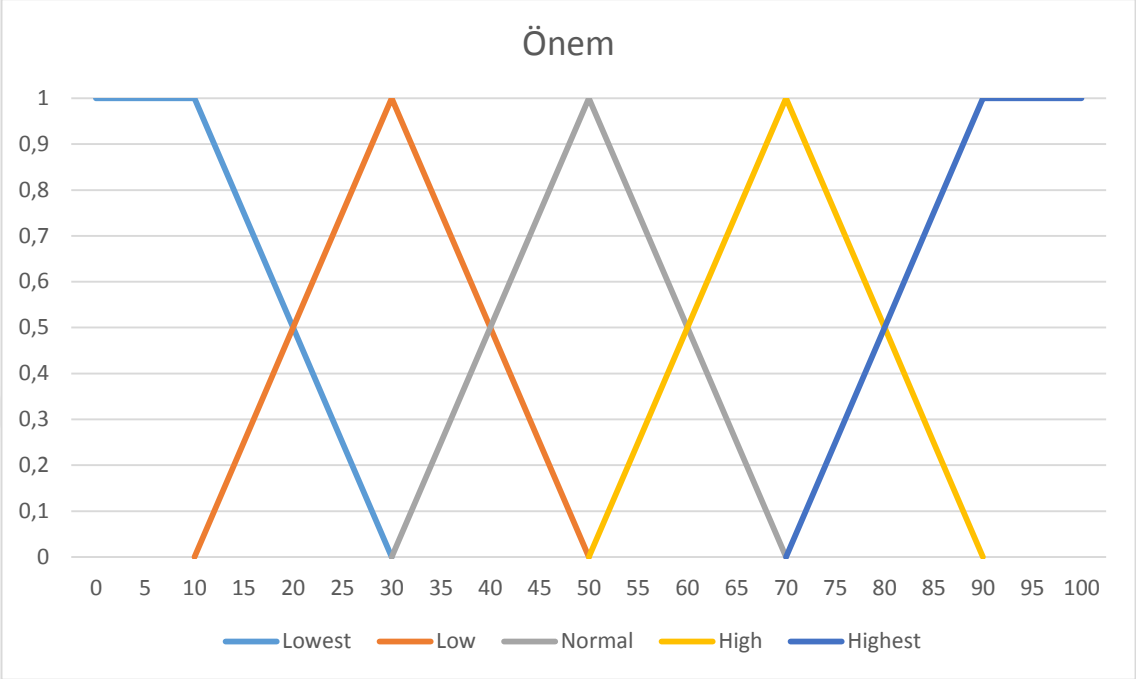
Z_j = j indeksindeki değer

4.2.1 Üyelik fonksiyonları

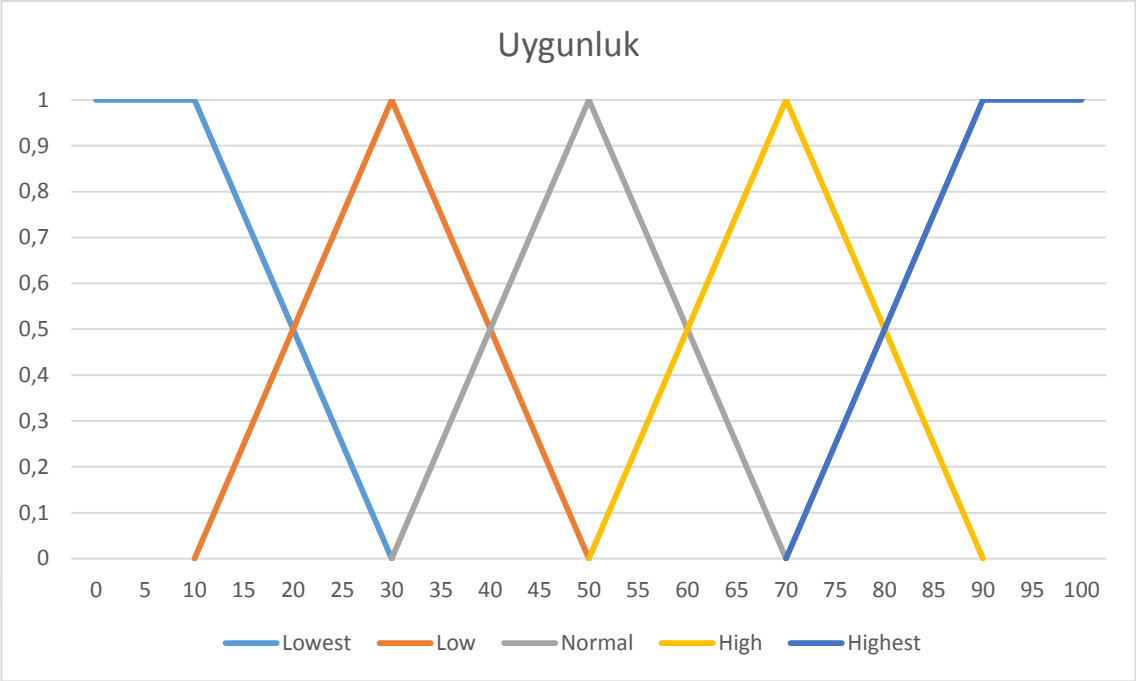
Bulanık mantıkta oluşturulan kümelerin sınırlarının, klasik mantığa göre daha belirsiz veya esnek olduğu ifade edilebilir. Klasik küme mantığında bir eleman herhangi bir kümeye net bir şekilde dâhildir veya değildir. Bulanık mantıkta ise kümelerin üyelik fonksiyonları vardır. Her elemanın bu fonksiyonlara göre bir veya birden fazla kümeye belli derecelerde dâhil olması söz konusu olabilir.

Örneğin, 1.70m boyunda olan bir kişi, orta boylular kümesine 0.80 oranında, uzun boylular kümesine ise 0.30 oranında dâhil olabilir. 1.85m boyunda olan bir kişi ise orta

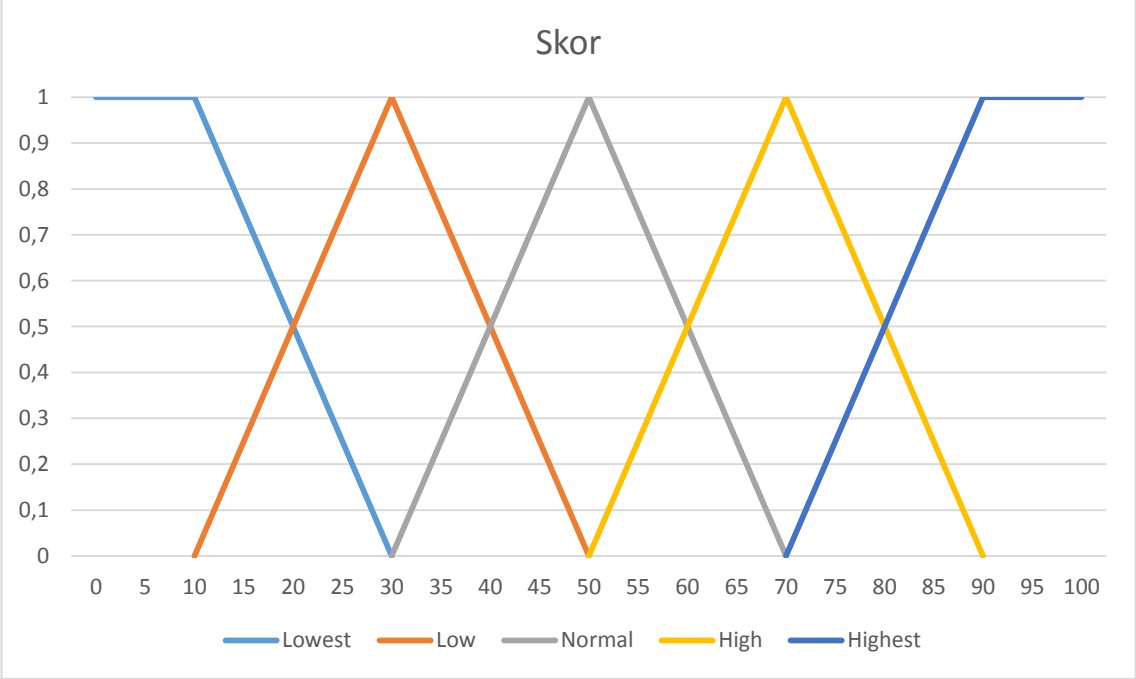
boylular kümesine 0.10 oranında, uzun boylular kümesine 0.95 oranında dâhil olabilir. Üyelik fonksiyonları bu çalışmada aşağıdaki şekilde tanımlanmıştır;



Şekil 4.4 Girdi Üyelik Fonksiyonu: Önem



Şekil 4.5 Girdi Üyelik Fonksiyonu: Uygunluk



Şekil 4.6 Çıktı Üyelik Fonksiyonu: Skor

Kullanıcılar klasik değerlendirme yöntemlerinde, belirli sayıda seçenekten birisini tercih etmek zorunda kalmaktadır. Kendisine sunulan "Düşük", "Orta", "Yüksek" gibi seçeneklerin arasından kendisine en yakın hissettiğini, bazen de iki seçenek arasında bir değer sunulmadığı için tercihinin altındaki veya üstündeki seçeneği rasgele seçmektedir. Bu çalışmada sürgü (slider) ölçeği kullanılarak kullanıcıların sezgisel olarak değerlendirme yapabilmesine imkân tanınmıştır. Sürgü ölçeğinde rakamsal değerler kullanılmamış olup, tamamen sürgü ölçeğinin sunduğu görsellik üzerinden, kullanıcılar sürgüyü dilediği kadar kaydırarak değerlendirmelerini sezgisel olarak yapabilmektedirler.

Klasik yönteme göre daha esnek bir yapı sunan bu yöntemde kullanıcılar dilediği puanı verebildiği için "Önem" ve "Uygunluk" girdileri sözel parametrelerle tanımlanmış 5 eşit aralığa ayrılmıştır. Sürgü ölçeği aracılığıyla verilen puanlar, tanımlanmış üyelik fonksiyonlarına göre bu aralıklarla ilişkilendirilir. Tek çıktı parametresi olan "Skor", girdi parametreleri gibi üyelik fonksiyonlarına göre 5 eşit aralığa ayrılmıştır. Her bir ilke için verilen "Önem" ve "Uygunluk" puanları tanımlanmış kural setine göre bulanık çıkarım yöntemi uygulanarak ilgili "Skor" çıktısına dönüştürülür.

Bu çalışmada, ileri seviye mühendislik uygulamaları kullanılmadığı için ve karmaşık bir problem olmaması nedeniyle üçgen üyelik fonksiyonu kullanılmıştır. Girdiler ve çıktı, Likert tipi ölçekte olduğu gibi toplamı 100 olacak şekilde 5 eşit parçaya bölünmüştür. Her fonksiyonun en büyük değerini aldığı nokta, yani orta değer noktası, bir önceki üçgen üyelik fonksiyonun bittiği ve bir sonraki fonksiyonun başladığı nokta olarak belirlenmiştir. Örneğin Şekil 4.6'te "Normal" olarak görünen üçgen üyelik fonksiyonunun asıl aralığı 40 ile 60'tır. Ancak ara değerlerin bulanıklaştırılması amacıyla üçgenin başlangıç noktası 30, bitiş noktası ise 70 olarak belirlenmiştir. İlk ve son fonksiyon dışında bütün fonksiyonlarda bu yöntem uygulanmıştır. İlk ve son fonksiyonda ise sırasıyla orta değerden öncesi ve orta değerden sonrası için fonksiyona mutlak üyelik olacak şekilde tasarlanmıştır.

4.2.2 Kurallar

Bulanıklaştırıcı vasıtasıyla 0 ile 1 arasında değerlerle üyelik dereceleri belirlenen girdiler, kural setinde tanımlı olan kurallara göre işlemden geçirilerek çıktı dağılımı kümesi hesaplanır. Tek girdisi olan kurallarda çıktının üyelik fonksiyonu, girdinin üyelik derecesine göre kırpılır. Birden fazla girdisi olan kuralların ise "ve", ya da "veya" operatörüne göre en küçük değeri (min) ya da en büyük değeri (max) alınarak kuralın ağırlığı hesaplanır.

Uzman görüşleri ile netleştirilen kural setinde, "Önem" düşük olan kullanılabilirlik ilkelerinde, "Uygunluk" girdisinin etkisinin daha az olması ve çıktı değeri olan "Skor" parametresinin değerinin küçük olması sağlandı. "Önem" yüksek olan ilkelerde ise bu parametrenin etkisinin daha belirgin olduğu bir kural seti oluşturuldu. Yani önemli ve uygunluğu en yüksek ilkeler, en yüksek puanları aldılar. Daha az önemli ve uygunluğu yüksek olan ilkeler ise daha düşük puanlar aldılar. Bir başka ifade ile önem derecesinin düşük veya yüksek olmasına göre ilkelerin alabileceği nihai skor değer aralığı daha küçük veya büyük oldu.

Aşağıda verilen kurallar bu durum için örnek teşkil edebilir;

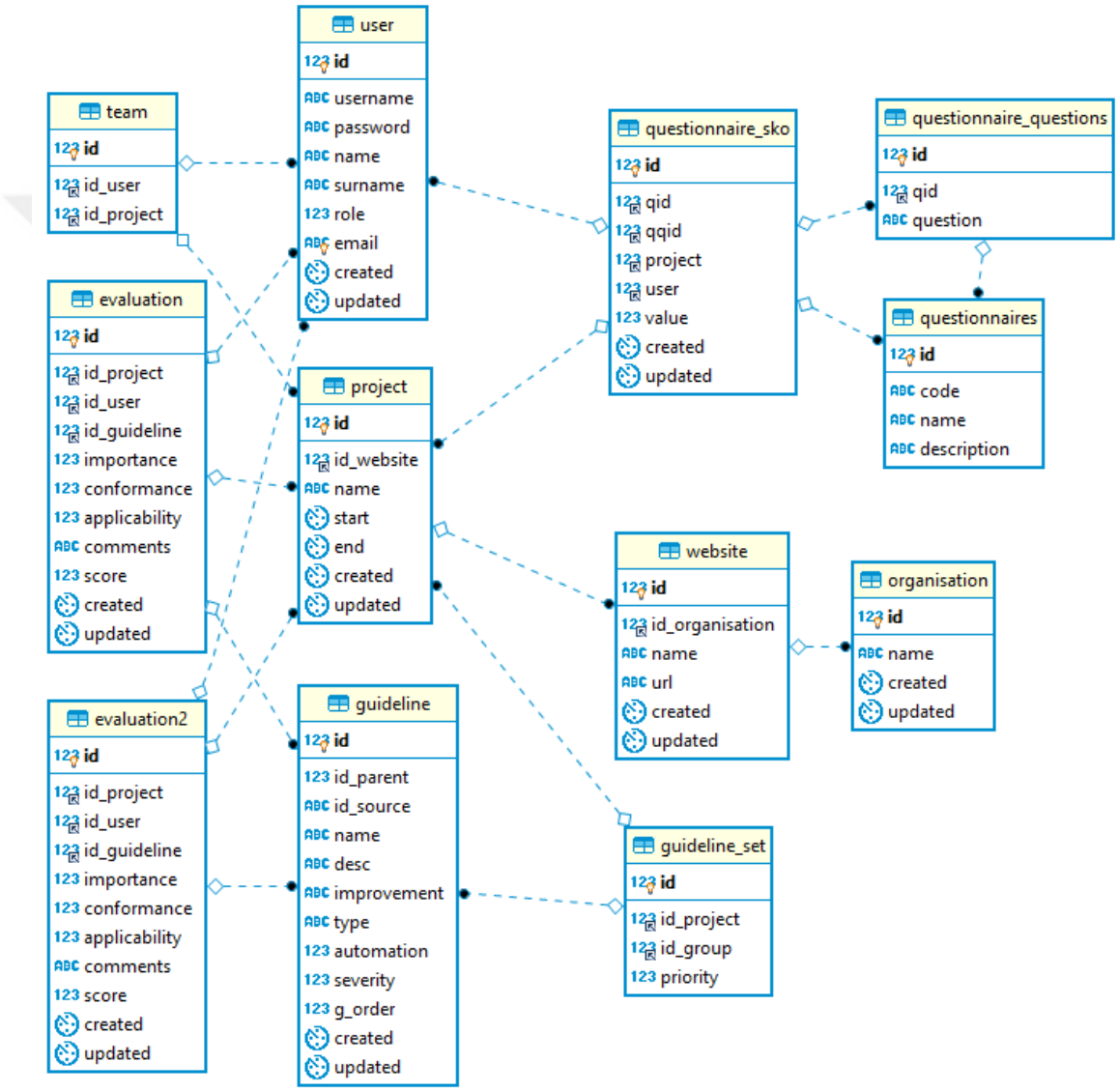
IF importance.lowest AND conformance.high THEN score.low
IF importance.low AND conformance.highest THEN score.normal
IF importance.normal AND conformance.highest THEN score.high
IF importance.high AND conformance.lowest THEN score.lowest
IF importance.highest AND conformance.lowest THEN score.lowest

4.3 Araç: Kullanılabilirlik Değerlendirme Sistemi (KDS)

İnternet sitelerinin kullanılabilirlik açısından değerlendirilebilmelerini kolaylaştırmak amacıyla PHP, MySQL, Javascript, Bootstrap gibi teknoloji ve kütüphaneler kullanılarak, uzman kullanıcıların ve son kullanıcıların internet sitelerini değerlendirme projelerinde yer alabileceği bir sistem tasarlanmıştır. Bu sisteme KAMİS Rehber

içerisinde yer verilen tüm kategori, ilke ve yönergeler eklenmiş, aynı zamanda Sistem Kullanılabilirlik Ölçeği (SKÖ) anketi de oluşturularak son kullanıcıların internet sitelerini 10 kısa soru üzerinden değerlendirebileceği bir yapı tasarlanmıştır.

KDS aracının veritabanı için ER diyagramı aşağıda verilmiştir;



Şekil 4.7 KDS aracı ER Diyagramı

Veritabanı tarafında toplam 12 tablo oluşturulmuştur. Kullanılabilirlik kategorileri, ilkeleri ve yönergeleri "guideline" tablosunda tutulmuştur. Likert ölçeğine göre

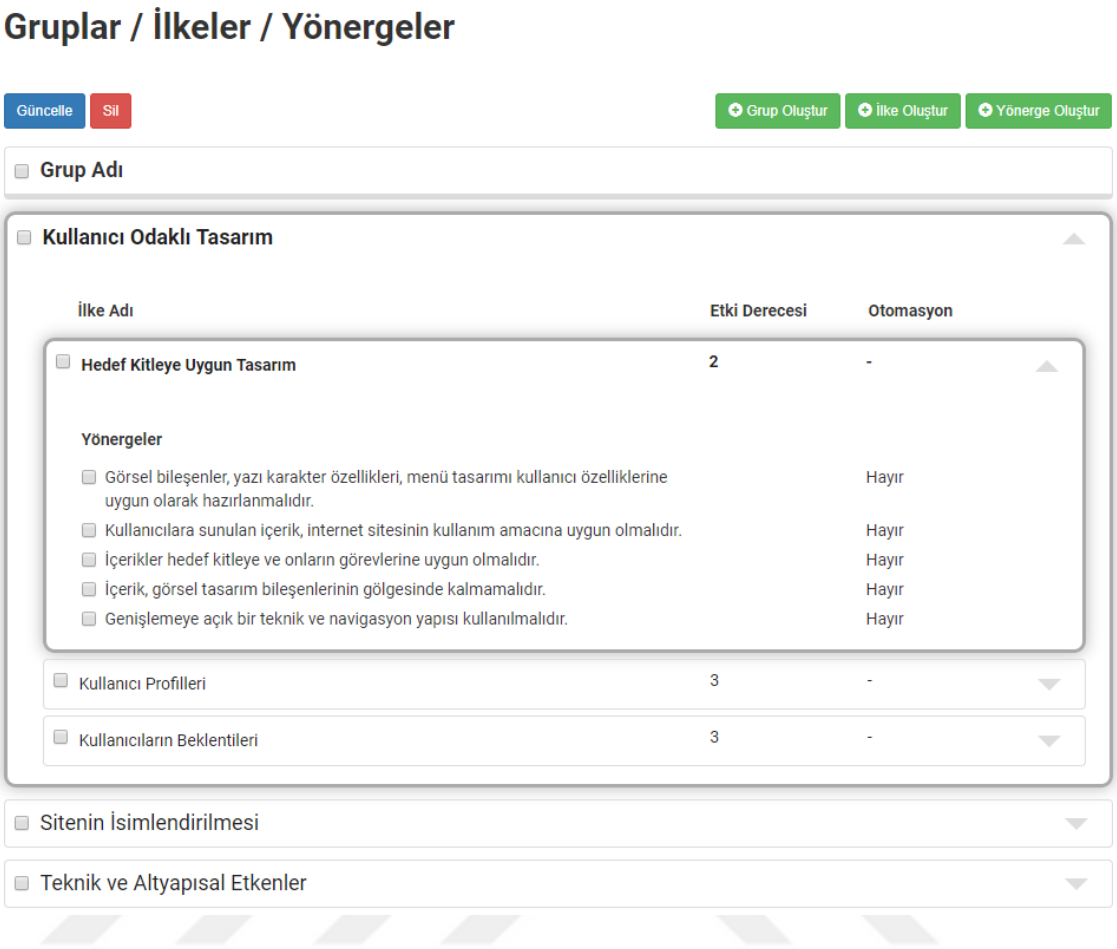
değerlendirme verileri “evaluation” tablosunda, sürgü ölçeğine göre değerlendirme verileri ise “evaluation2” tablosunda tutulmuştur.

Sistem üzerinden gerçekleştirilebilen işlemler aşağıda sıralanmıştır;

1. İlke seti yönetimi
2. Kullanıcı yönetimi
3. Kurum ve internet sitesi yönetimi
4. Proje yönetimi
5. Proje değerlendirme ekranları
    a. Uzmanlar için klasik değerlendirme
    b. Uzmanlar için bulanık değerlendirme
    c. Son kullanıcılar için anket değerlendirmeleri
6. Proje değerlendirmelerine ait raporların görüntülenmesi

4.3.1 İlke seti yönetimi ekranı

Bu bölümde yer alan ekranlar, herhangi bir ilke setinin eklenmesini mümkün kılmaktadır. Ancak çalışma kapsamında KAMİS Rehber referans alındığı için bu rehberde yer alan ilkeler ve yönergeler eklenerek çalışma gerçekleştirilmiştir.



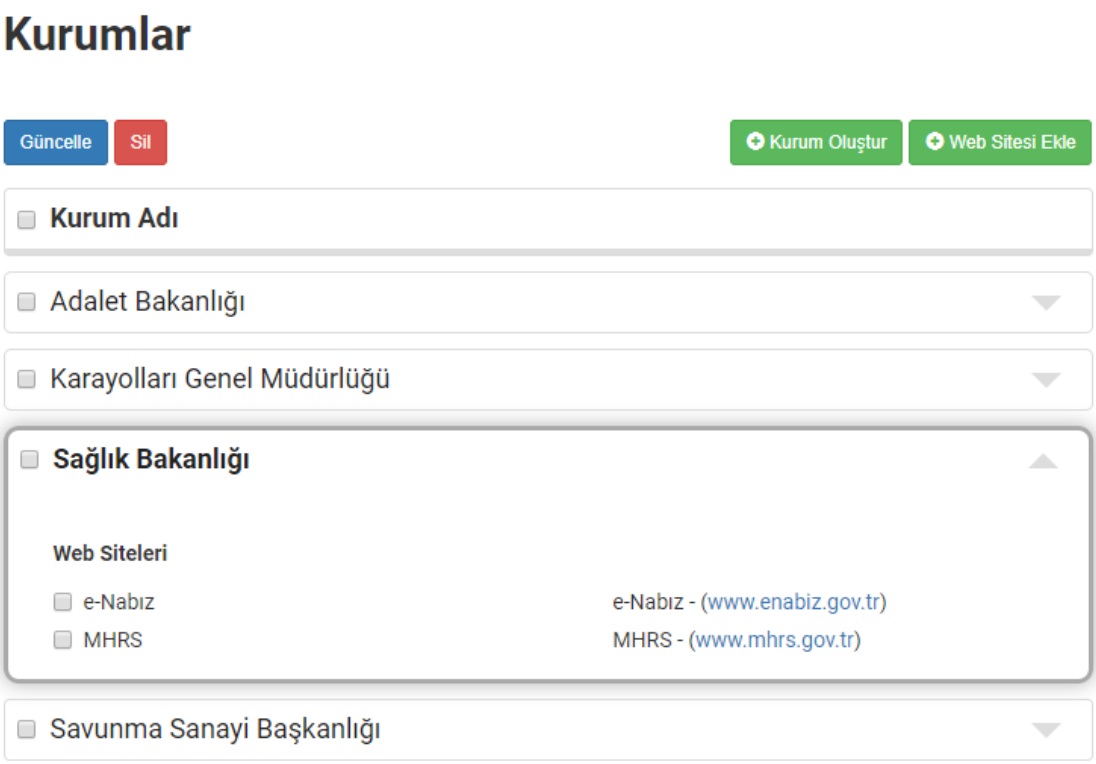
Şekil 4.8 İlke Seti Yönetimi Ekranı

4.3.2 Kullanıcı yönetimi ekranı

Kullanıcı yönetimi bölümünde KAMİS Rehber ilkelerine göre değerlendirme yapacak olan uzman kullanıcılar ve SKÖ anketini değerlendirecek olan son kullanıcılar gibi farklı türde kullanıcıların eklenmesi, güncellenmesi ve silinmesi gibi işlemler gerçekleştirilebilmektedir.

4.3.3 Kurum ve internet sitesi yönetimi ekranı

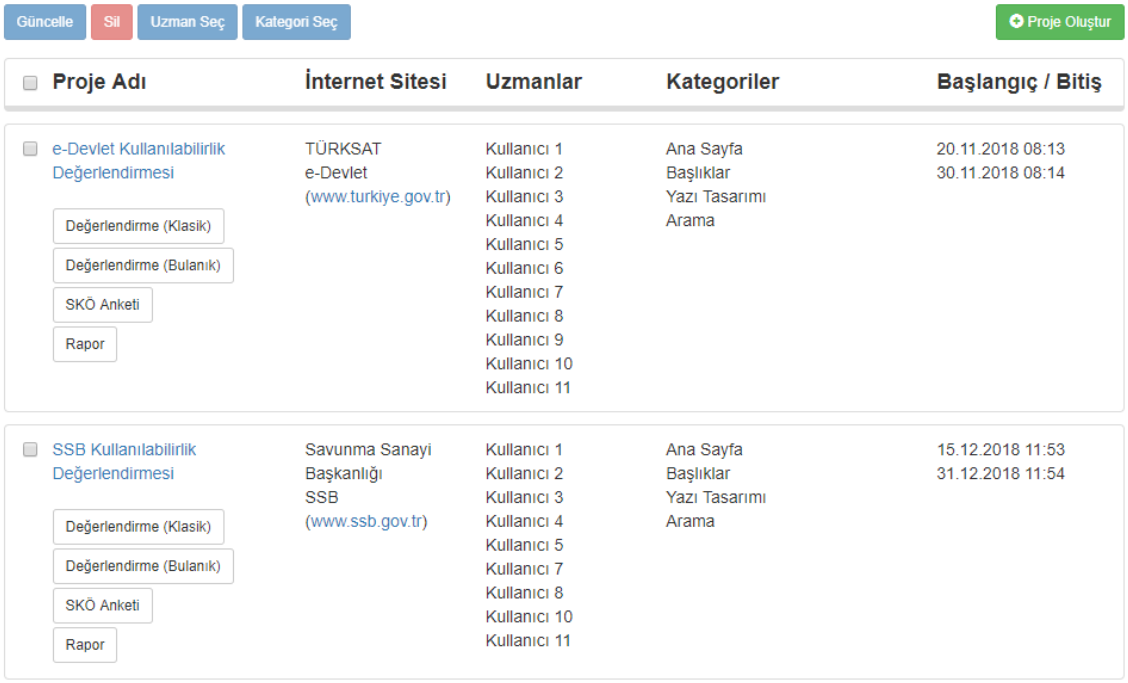
Kurum yönetiminin yapıldığı bu ekranlarda her kurum için kullanılabilirlik değerlendirmesi yapılacak internet siteleri de eklenebilmektedir.



Şekil 4.9 Kurum ve İnternet Sitesi Yönetimi Ekranı

4.3.4 Proje yönetimi ekranı

Sistemde, kurum ve ona ait bir internet sitesi seçilerek proje oluşturulabilmektedir. Daha sonra, bu oluşturulan projenin KAMİS Rehber'de yer alan hangi kategoriler açısından değerlendirileceği ve hangi uzmanların bu değerlendirmeye katılacağı seçilebilmektedir.



Şekil 4.10 Proje Yönetimi Ekranı

4.3.5 Proje değerlendirme ekranları

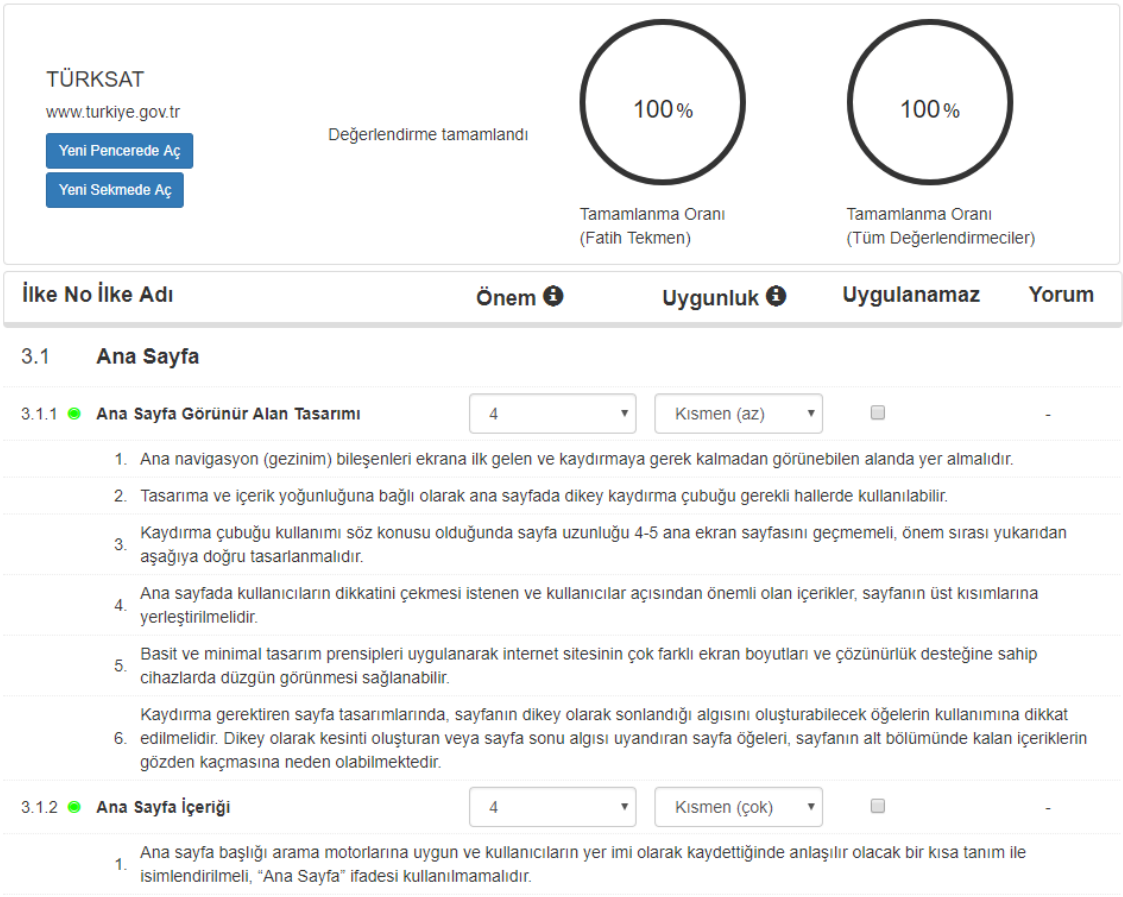
Proje değerlendirmeleri farklı ekranlar üzerinden yapılmaktadır. Uzman kullanıcılar, KAMİS Rehber ilkelerini klasik ve bulanık değerlendirme ekranlarından görüntüleyerek puanlayabilmektedir. Aynı zamanda uzman kullanıcılar ve son kullanıcılar SKÖ anketi üzerinden değerlendirme yapabilmektedir.

TS EN ISO 9241-151 standardına ait kontrol listesinde "Uygulanabilirlik" ve "Uygunluk" parametreleri üzerinden değerlendirme yapılmaktadır. KDS aracında uzman kullanıcıların her ilkenin uygulanabilirliğini girebilmesi için "Uygulanamaz" başlıklı bir alan tanımlanmış, buradan seçim kutusunun işaretlenmesine ve ilkenin uygulanabilirliğinin değerlendirilmesine imkân tanınmıştır. "Uygunluk" başlıklı alandan, üzerinde çalışılan internet sitesinin ilkeye uygunluğu değerlendirilmiştir. Kontrol listesinde yer almayan, ancak KDS aracında yer verilen "Önem" başlıklı alan vasıtasıyla uzman kullanıcıların her ilkenin önem derecesini ataması ve uzman görüşüne göre ilkelerin ağırlıklandırılması sağlanmıştır.

4.3.5.1 Likert tipi klasik değerlendirme ekranı

Bu ekranda kullanıcılar internet sitesini rehber ilkeleri üzerinden açılır kutucukların içerisinde yer alan seçenekleri kullanarak değerlendirebilirler. "Önem" açılır kutucuğundaki değerler 1 ile 5 arasında değişmektedir. "Uygunluk" açılır kutucuğunda ise "Uygun değil", "Kısmen (az)", "Kısmen (orta)", "Kısmen (çok)" ve "Uygun" olmak üzere 5 farklı seçenek vardır. "Uygunluk" değerlerinin rakamsal karşılığı ise 0 ile 4 arasında değişmektedir. "Uygun değil" seçeneği tercih edildiğinde rakamsal olarak arkaplanda 0 değeri ile işlem yapılmaktadır.

Aynı zamanda ilkenin internet sitesi açısından değerlendirilmesi uygun değilse "Uygulanamaz" seçim kutusu işaretlenebilmekte ve ilke ile ilgili yorumlar yazılabilmektedir.



Şekil 4.11 Klasik Değerlendirme Ekranı

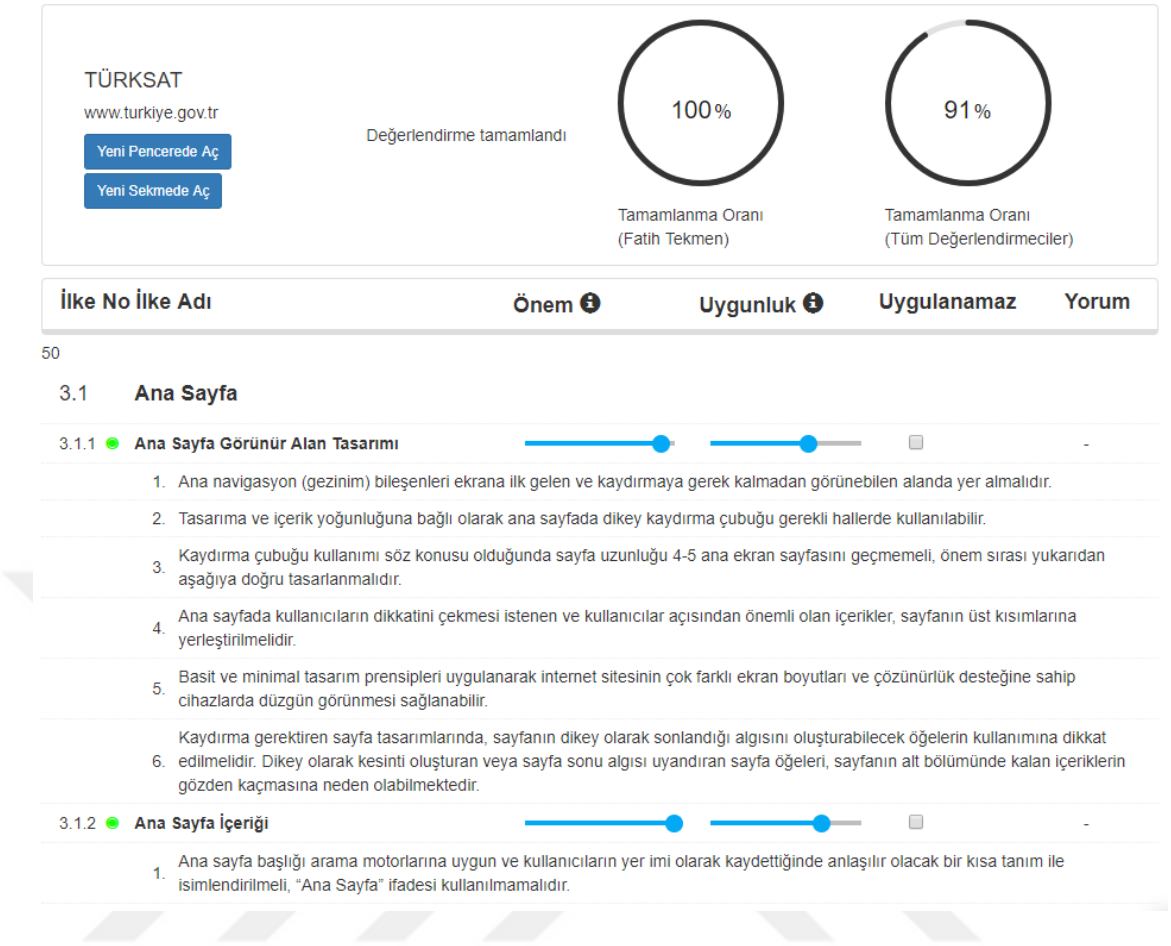
Burada kullanıcılar tarafından değerlendirilen her ilke için “Önem” ve “Uygunluk” değerleri kendi arasında çarpılmakta ve sonucun 0 ile 100 arasında bir değere dönüştürülebilmesi için tekrar 5 ile çarpılmaktadır. “Önem” alanı her ilke için ağırlık değeri olarak görev yapmaktadır. “Uygunluk” alanından “Uygun değil” seçildiğinde ise bu ilkenin önemi yüksek dahi olsa, skorunu sıfırlamaktadır.

4.3.5.2 Sürgü ölçeği değerlendirme ekranı

Bulanık değerlendirme ekranının klasik değerlendirmeden tek farkı, değerlendirmelerin bir sürgü ölçeği yardımı ile yapılabiliyor olmasıdır. Kullanıcıların daha özgür değerlendirme yapabilmesi ve ilkeleri daha sezgisel değerlendirebilmeleri için sürgü ölçeği üzerinden de çalışabilmelerine imkân tanınmıştır. Sürgü kaydırılırken kullanıcılara herhangi bir rakamsal değer gösterilmemekte, kullanıcılar sadece sürgü ölçeğinin görselliği üzerinden sezgisel olarak puanlamalarını yapabilmektedirler.

İnsanların verdikleri kararlarda o anki ruh hallerinin etkisi üzerine araştırmalar yapılmıştır. Literatür bölümünde bahsedilen çalışmaya göre (de Vries vd. 2008) insanların verdikleri karar, mutlu oldukları zamanlarda daha hızlı ve sezgisel olmakta, üzgün ruh hallerinde ise kararlarını analiz etmek için daha fazla düşündükleri, artıları ve eksileri daha fazla değerlendirdikleri görülmüştür. Bulanık değerlendirme ekranında bu durumun etkilerini azaltmak için hem sürgü ölçeği kullanılarak, kullanıcıların daha rahat ve sezgisel bir değerlendirme yapabilmesi sağlanmış, hem de verilen kararların bulanık çıkarım yöntemi ile kümelenmesi ve uzman görüşleri ile şekillenen kural setine göre doğruya yakınsanması amaçlanmıştır.

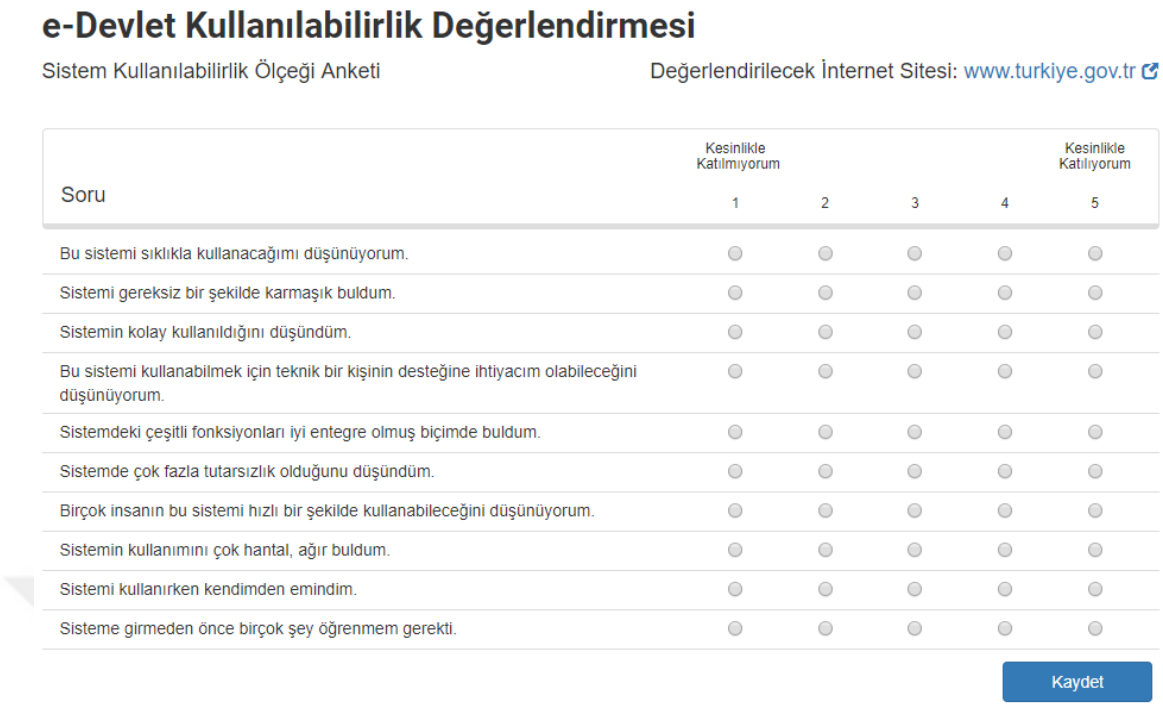
Sürgü ölçekleri görsel ve sezgisel olarak kullanıcılara geribildirim vermektedir. Sürgü ölçeklerinde küçük veya büyük rakamsal değerlerin kullanılması, bu sistemi kullanan kişilerin algısını veya değerlendirme sonuçlarını etkileyebileceği (Bagchi ve Davis 2016) için kullanıcılar sürgüyü kaydırdıklarında herhangi bir rakamsal değer gösterilmemektedir.



Şekil 4.12 Bulanık Değerlendirme Ekranı

4.3.5.3 Sistem kullanılabilirlik ölçeği anketi ekranı

Uzman kullanıcılar ve son kullanıcılar bu ekran üzerinden ilgili proje için anket değerlendirmesi yapabilmektedir. SKÖ anketinin toplam 10 tane, kısa, anlaşılır ve net sorusu üzerinden değerlendirmeler yapılmaktadır.



Şekil 4.13 Sistem Kullanılabilirlik Ölçeği Anketi Ekranı

SKÖ anketinde olumlu sorulara verilen puanlardan 1 çıkarılır. Olumsuz sorulara verilen puanlar ise 5'ten çıkarılır. Böylelikle toplam 10 sorudan 0 ile 40 arasında bir puan elde edilmekte, bunun 0 ile 100 arasında bir değere dönüştürülmesi için de bu puan 2,5 ile çarpılmaktadır.

Kullanıcılar tarafından yapılan tüm değerlendirmeler sonucunda 0 ile 100 arasında bir puan üretilerek tüm bu değerlendirmelerin birbiri ile karşılaştırılabilir olması sağlanmıştır.

4.3.6 Rapor ekranları

Proje değerlendirmelerine ilişkin raporlar bu ekranlarda görüntülenebilmektedir. Rapor ekranında sırası ile aşağıdaki raporlar görülebilir;

- Kategori bazında klasik ve bulanık uzman değerlendirmesine ait radar grafiği
- Bulanık uzman değerlendirmesi ve SKÖ anketi ortalamaları bar grafiği

- Uzman ve son kullanıcılardan SKÖ anketi ortalamaları için bar grafiği
- Kategori bazında tüm uzman kullanıcıların klasik değerlendirme bar grafiği
- Kategori bazında tüm uzman kullanıcıların bulanık değerlendirme bar grafiği
- Kategori bazında tüm uzman kullanıcıların klasik ve bulanık değerlendirmelerinin yan yana görüntülendiği bar grafiği
- İlke bazında tüm uzman kullanıcıların klasik değerlendirme çizgi grafiği
- İlke bazında tüm uzman kullanıcıların bulanık değerlendirme çizgi grafiği

Rapor ekranları, yapılan değerlendirmeler ve çalışmalar üzerinden üretilen verinin anlaşılabilmesi, analiz edilebilmesi açısından önemlidir. Bundan dolayı bu ekran üzerinden mümkün olduğunca fazla veri gösterilmeye çalışılmıştır.

4.4 Bölüm Özeti

İlkeler üzerinden yapılan kullanılabilirlik çalışmalarında, değerlendiricilerin öznel görüşleri ile değerlendirme puanlarının ortaya çıkarılması, özgür değerlendirme yapamaması ve yalnızca kullanılabilirlik alanında tecrübeli kişilerle bu tür çalışmaların yapılabilmesi gibi problemler, bu tez kapsamında dikkate alınmış ve buna yönelik çözüm önerisinde bulunulmuştur. Bu kapsamda tarayıcı üzerinde çalışan bir sistem geliştirilmiş ve kullanıcıların Likert tipi ölçekle, sürgü ölçeğiyle ve SKÖ anketi ile 3 ayrı değerlendirme yapması sağlanmıştır. Değerlendirmelerdeki öznelliği ve kullanılabilirlik alanında uzmanlığı olan bir kullanıcı ihtiyacını ortadan kaldırmak için her ilke altında detaylı yönergelerin olduğu KAMİS Rehber ilke setleri üzerinden değerlendirme yapılmıştır.

Çalışma kapsamında kullanıcıların sırasıyla bu üç farklı değerlendirmeyi yapması sağlanmış ve Likert tipi ölçekle ve bulanık çıkarım sistemi ile desteklenen sürgü ölçeğinde elde edilen puanlarla SKÖ anketi puanı karşılaştırılmıştır. Literatürde tanımlı olan ve günümüzde de yaygın kullanılan SKÖ anket yöntemi referans alınarak her iki ölçek tipinde ortaya çıkan puanların SKÖ anketinden elde edilen ortalama puana yakınlığı değerlendirilmiştir.

Önerilen yöntemin, teknik uzmanlığı olan kullanıcıların hızlı ve düşük maliyetli olarak kullanılabilirlik değerlendirmeleri yapabilmeleri, objektif ve özgürce değerlendirebilmeleri ve uzman / son kullanıcı değerlendirmeleri arasındaki farkın görülebilmesi açısından katkı sağlaması hedeflenmiştir. Uzman ve son kullanıcıların yaptıkları değerlendirmelerde çok farklı sonuçlar çıkabilmektedir. Standartlar, sezgiseller ve ilkeler üzerinden yapılan değerlendirmeler genelde uzman görüşünü yansıtmakta; kullanıcı testleri ise kullanıcıların algısını ve bakış açısını ortaya koyan testler olarak kullanılmaktadır. Kullanıcı testleri, yüksek maliyetli ve uzun süreli çalışmalar olmaktadır. Bu tez çalışmasında, hem uzman görüşünün hem de son kullanıcı algısının hızlı bir şekilde ölçülebildiği bir yöntem kullanılmış ve çıkan sonuçlar karşılaştırılmıştır.

5. BULGULAR ve TARTIŞMA

KDS aracı sayesinde, internet sitelerine ait kullanılabilirlik değerlendirmeleri, bu alanda doğrudan uzmanlığı olmayan yazılım mühendisi, test mühendisi, tasarımcı gibi kişiler tarafından kolaylıkla yapılabilir hale gelmiştir. KAMİS Rehber içerisinde yer alan ve her ilke için yazılan detaylı yönergeler, bu sistem içerisinde de tanımlanmış ve kullanıcıların değerlendirmelerini yapmalarında yardımcı unsur olarak etkili olmuştur.

Toplamda 3 değerlendirme projesi oluşturulmuş ve kullanıcılardan bu projeler kapsamında, farklı e-devlet internet sitelerini değerlendirmeleri istenmiştir. İlk internet sitesi için yapılan değerlendirme de kullanıcıların verdiği puanlara müdahale edilmemiştir. Ancak diğer projelerde kullanıcılar değerlendirmelerini yaptıktan sonra aykırı değerlere sahip ilkeler, değerlendirmeyi yapan kullanıcılara tekrar sorulmuş ve bu ilkelerin kullanıcıların bir kısmı tarafından farklı algılandığı anlaşılmıştır. Aykırı değerlere sahip bu ilkeler kullanıcılar tarafından tekrar değerlendirilmiştir.

İlk proje kendi içerisinde daha detaylı olarak değerlendirilmiş olup bu kapsamda klasik ve bulanık değerlendirmelerin karşılaştırılması, bulanık değerlendirme ile SKÖ anketi sonuçlarının karşılaştırılması, uzman kullanıcı ve son kullanıcı SKÖ anketi sonuçlarının karşılaştırılması gibi alanlarda incelenmiştir. Diğer iki proje ise bulanık değerlendirme açısından sistem tutarlılığının ölçülmesi ve güvenilirliğin artırılması amacıyla kullanılmıştır.

Bulanık çıkarım yöntemi, Likert tipi ölçeğin kullanıldığı değerlendirme yönteminde de kullanılabilirdi. Ancak burada klasik yönteme alternatif olarak sürgü ölçeği ile kullanıcıların daha sezgisel değerlendirme yapabilmelerine imkân tanınmış ve bu ikinci yöntem bulanık çıkarım sistemi ile desteklenerek daha doğru sonuçlar elde edilmiş oldu. Likert tipi ölçek ile yapılan çalışmada kullanıcılar kararlarını ön tanımlı değerleri seçerek girdiği için burada, kullanıcıların kararını daha esnek verebilmesi amacıyla sürgü ölçeği yöntemi kullanılmıştır.

5.1 Likert Tipi ve Sürgü Ölçeği Değerlendirme Sonuçları

Bu çalışmada toplam 3 internet sitesi kullanılabilirlik açısından değerlendirildi. Ancak en detaylı değerlendirme çalışması, sadece ilk projede değerlendirilmek üzere seçilen www.turkiye.gov.tr sitesi üzerinde gerçekleştirildi. Bu projede uzman kullanıcılardan seçilen internet sitesini klasik yöntemle ve bulanık yöntemle değerlendirmeleri istendi.

Klasik yöntemde kullanıcılar, her bir kullanılabilirlik ilkesi için önem ve uygunluk seçim kutularından ön tanımlı 5 farklı değer arasından seçim yaparak değerlendirmelerini tamamladılar. Bulanık yöntemde ise seçim kutularının yerine sürgü ölçeği kullanıldı. Bu alanda kullanıcıların, herhangi sayısal bir değer gösterilmeden, sadece görsel olarak kaydırabildikleri bir çubuk üzerinden sezgisel olarak değerlendirmelerini yapmaları sağlandı.

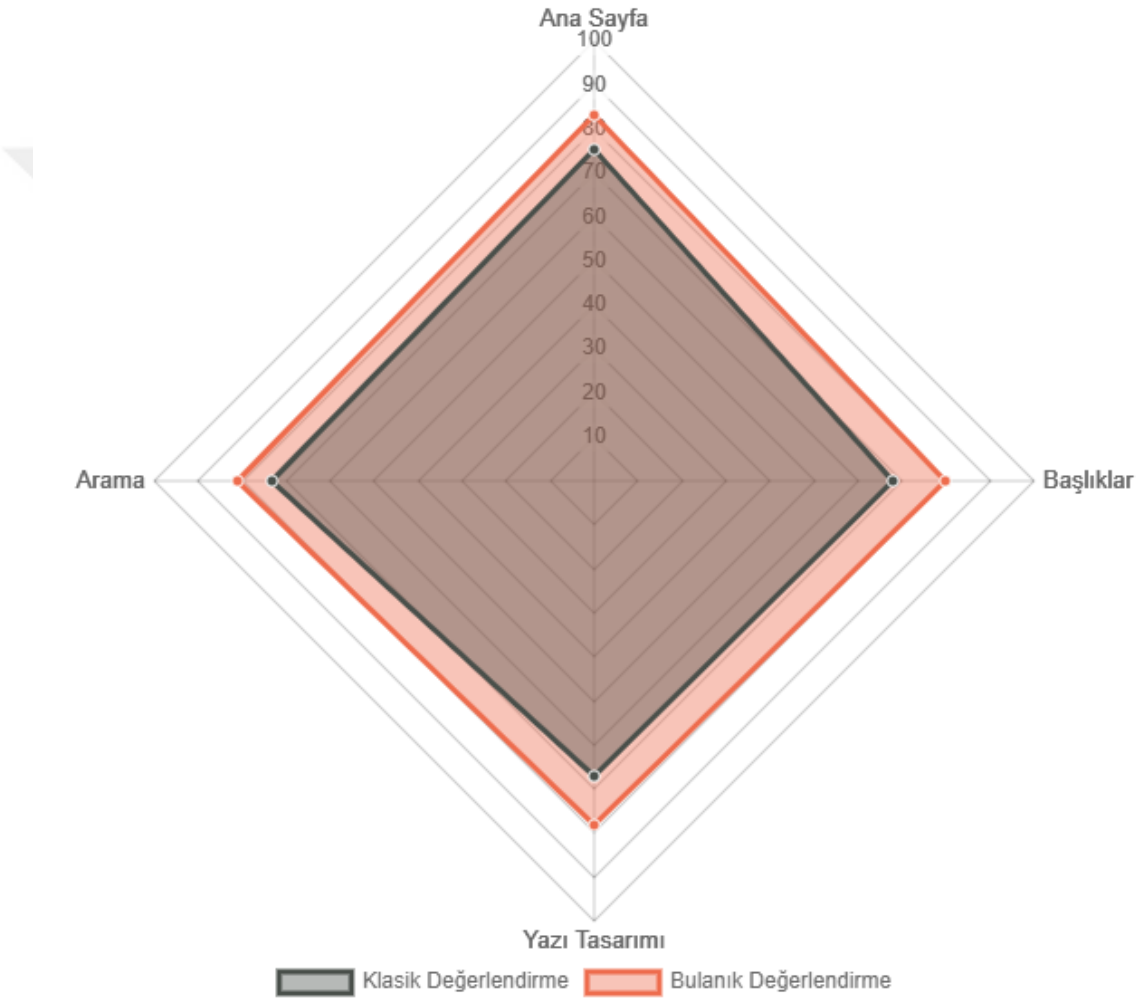
Klasik yöntemde, önem alanı için kullanıcılara 1 ile 5 arasında değişen değerleri seçmelerine imkân tanındı. Uygunluk alanı içinse 0 ile 4 arasında değişen değerlere denk gelen ifadelerin seçilmesi mümkün kılındı. Her bir ilkenin puanı hesaplanırken kullanıcıların bu iki alandan seçtikleri değerler çarpılarak, bulanık değerlendirme ve SKÖ anketinden çıkan sonuçlarla kıyaslanabilmesi için 100 üzerinden bir değere dönüştürülebilmesi için 5 ile çarpıldı. Önem alanında seçilen değer, puan hesaplamasında ağırlıklandırma olarak değerlendirildi. Dolayısıyla ortalama puanın hesaplanmasında, burada ağırlıklı ortalama yönteminin kullanıldığı söylenebilir.

Bulanık yöntemde ise sürgü ölçeğinde sayısal bir değer gösterilmemesine rağmen kullanıcıların kaydırdıkları miktar, 100 üzerinden bir değere karşılık gelecek şekilde hesaplama yapıldı. Kullanıcılar, klasik yöntemde olduğu gibi çok daha kısıtlı değerlerden birisini seçmek zorunda kalmadan, daha sezgisel olarak değerlendirme yapma şansına sahip oldu. Aynı zamanda burada yapılan değerlendirmenin doğruluğunu artırmak amacıyla bulanık hesaplama yönteminden de faydalanılarak, kullanıcı puanlarının uzman görüşleri ile belirlenen bulanık kural setlerine tabi tutulması ve bir seviye daha işlemden geçirilmesi sağlandı. Çalışma sonucunda elde edilen veriler aşağıdaki adresten görüntülenebilir;

https://drive.google.com/open?id=1eGuJXtkkvtLEjJMkZwQGQwWSxptBNR6M

Kısa URL: http://bit.ly/2v3guO0

Klasik ve bulanık değerlendirme sonucunda, kategori bazında ortalamalar hesaplanarak radar grafik üzerinde gösterilmesi sağlandı. Aşağıda yer alan grafik elde edildi;



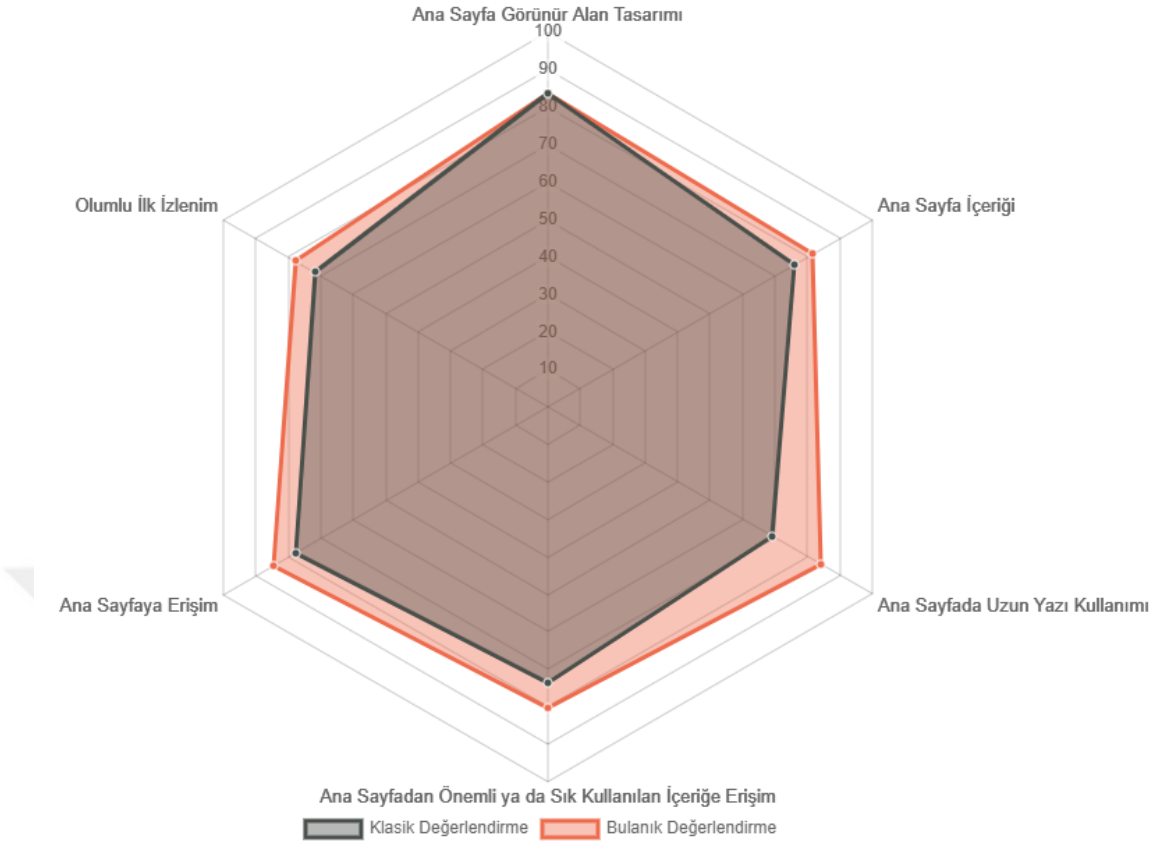
Şekil 5.1 Klasik ve Bulanık Değerlendirme Kategori Ortalamasına göre Radar Grafiği

Çizelge 5.1 Kategori Ortalamasına Göre Klasik ve Bulanık Değerlendirme Sonuçları

	Değerlendirme	
Kategori	**Klasik**	**Bulanık**
Ana Sayfa	75,30	81,97
Başlıklar	67,73	78,83
Yazı Tasarımı	67,27	76,27
Arama	73,07	79,69
Ortalama	**70,84**	**79,19**

Buradaki grafiğe göre bulanık değerlendirme sonuçlarından, klasik değerlendirmeye göre her kategoride daha yüksek ortalama puanlar elde edildiği görülebiliyor. Bulanık yöntemde, klasik yöntemden farklı olarak kullanıcıların daha özgür değerlendirme yapabilmesine imkân tanınmış olması ve aynı zamanda uzman görüşüne göre kullanıcıların verdiği puanların bulanık kural setlerine tabi tutularak işlemden geçirilmiş olması, daha yüksek ve doğru puanlamalar elde edilmesine yardımcı olmuştur.

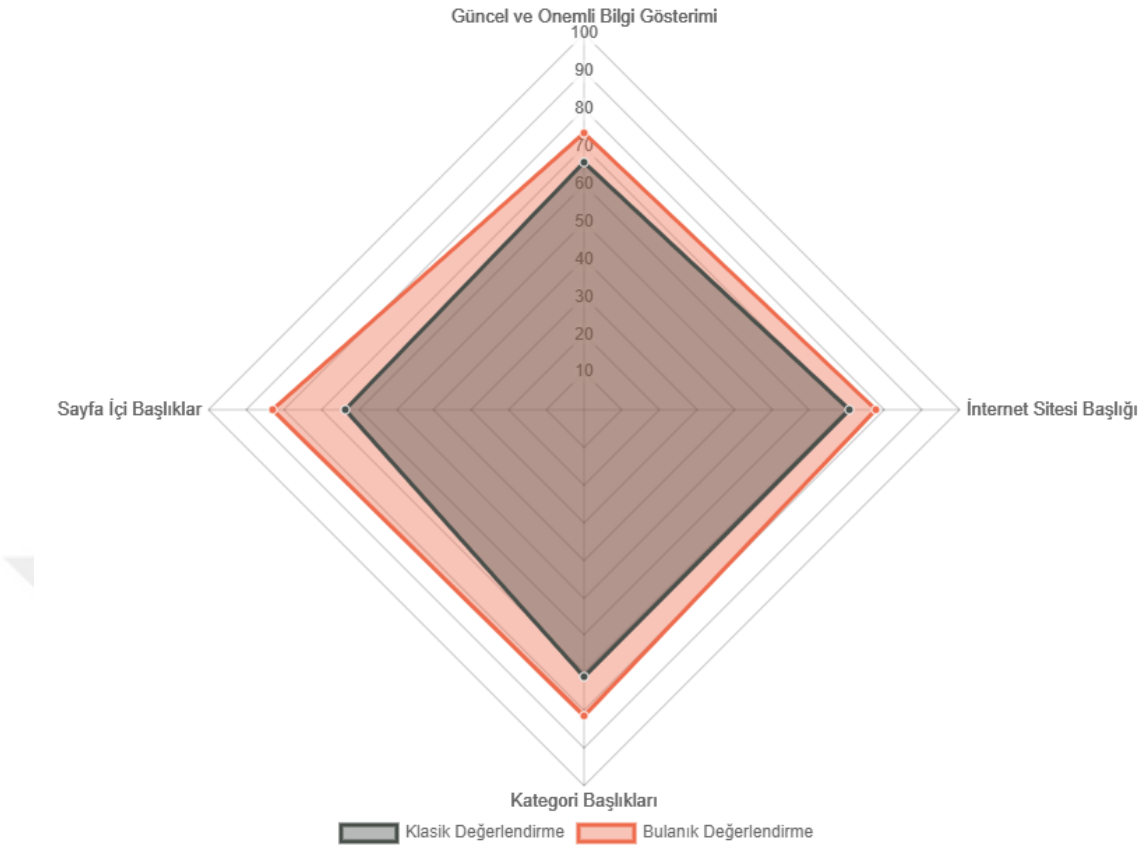
Genel değerlendirmeyi gösteren bu radar grafiğin dışında ayrıca değerlendirilmeye tabi tutulan bu 4 kategorinin her biri için klasik ve bulanık değerlendirme farkını gösteren ayrı radar grafikleri oluşturuldu. Bu grafikler aşağıda sırasıyla paylaşılmıştır;



Şekil 5.2 Anasayfa Kategorisine Ait İlkeler için Klasik ve Bulanık Değerlendirme Radar Grafiği

Çizelge 5.2 Ana Sayfa Kategorisi, Klasik ve Bulanık Değerlendirme Sonuçları

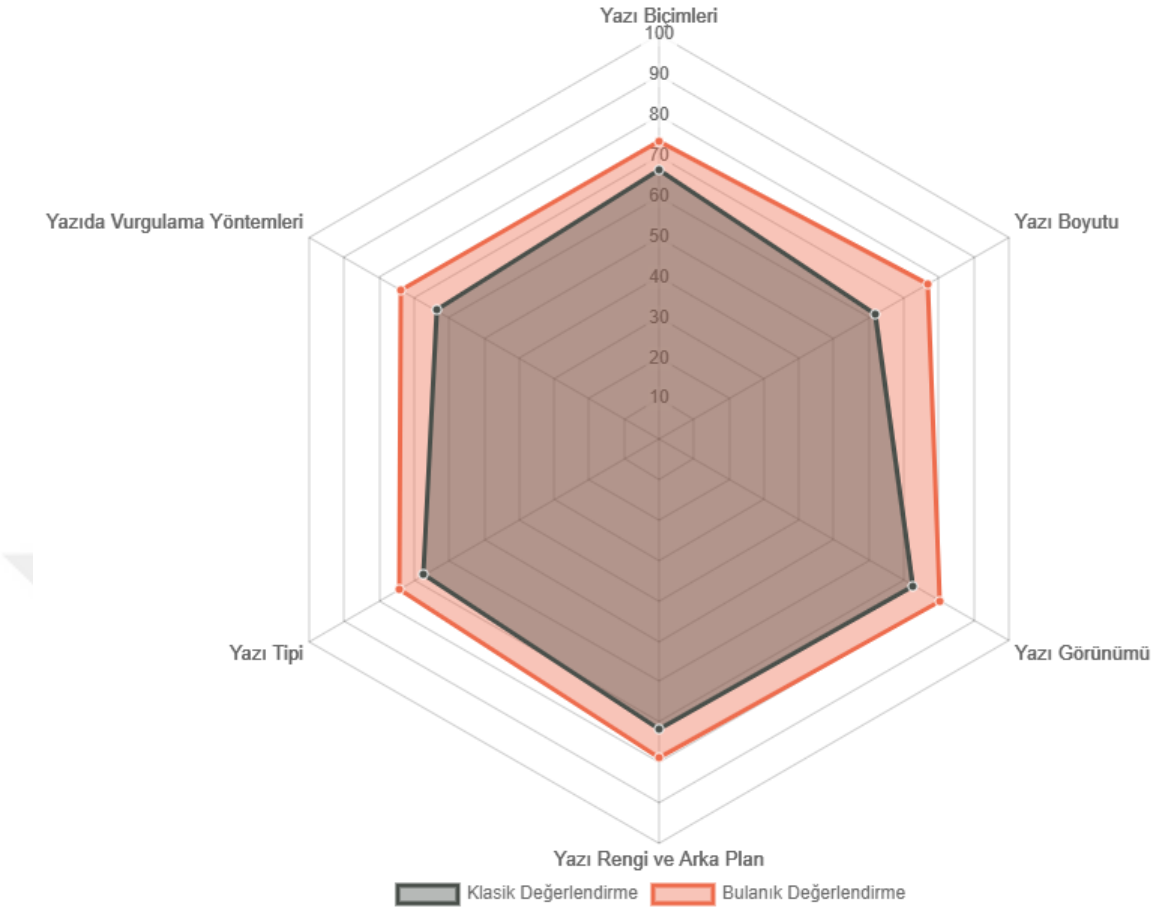
	Ana Sayfa Kategorisi / Değerlendirme	
İlke	**Klasik**	**Bulanık**
Ana Sayfa Görünür Alan Tasarımı	83,64	83,59
Ana Sayfa İçeriği	75,91	81,76
Ana Sayfada Uzun Yazı Kullanımı	69,09	84,04
Ana Sayfadan Önemli ya da Sık Kullanılan İçeriğe Erişim	73,64	80,18
Ana Sayfaya Erişim	77,73	84,54
Olumlu İlk İzlenim	71,82	77,71



Şekil 5.3 Başlıklar Kategorisine Ait İlkeler için Klasik ve Bulanık Değerlendirme Radar Grafiği

Çizelge 5.3 Başlıklar Kategorisi, Klasik ve Bulanık Değerlendirme Sonuçları

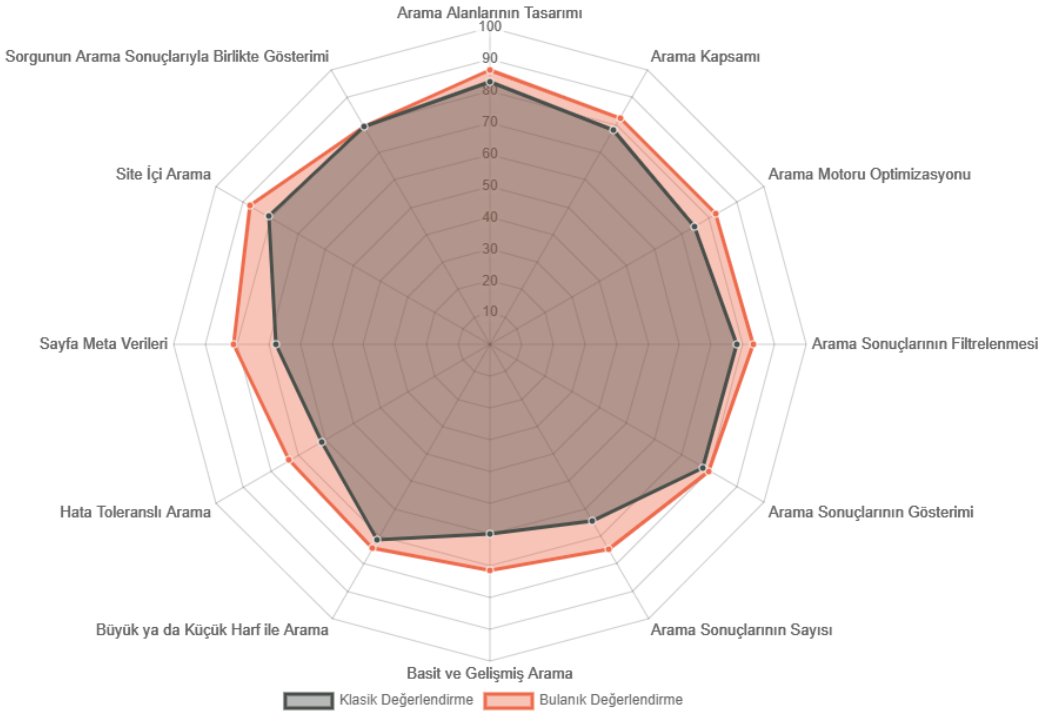
	Başlıklar Kategorisi / Değerlendirme	
İlke	**Klasik**	**Bulanık**
Güncel ve Önemli Bilgi Gösterimi	65,91	73,52
İnternet Sitesi Başlığı	70,45	77,55
Kategori Başlıkları	70,91	81,44
Sayfa İçi Başlıklar	63,64	82,81



Şekil 5.4 Yazı Tasarımı Kategorisine Ait İlkeler için Klasik ve Bulanık Değerlendirme Radar Grafiği

Çizelge 5.4 Yazı Tasarımı Kategorisi, Klasik ve Bulanık Değerlendirme Sonuçları

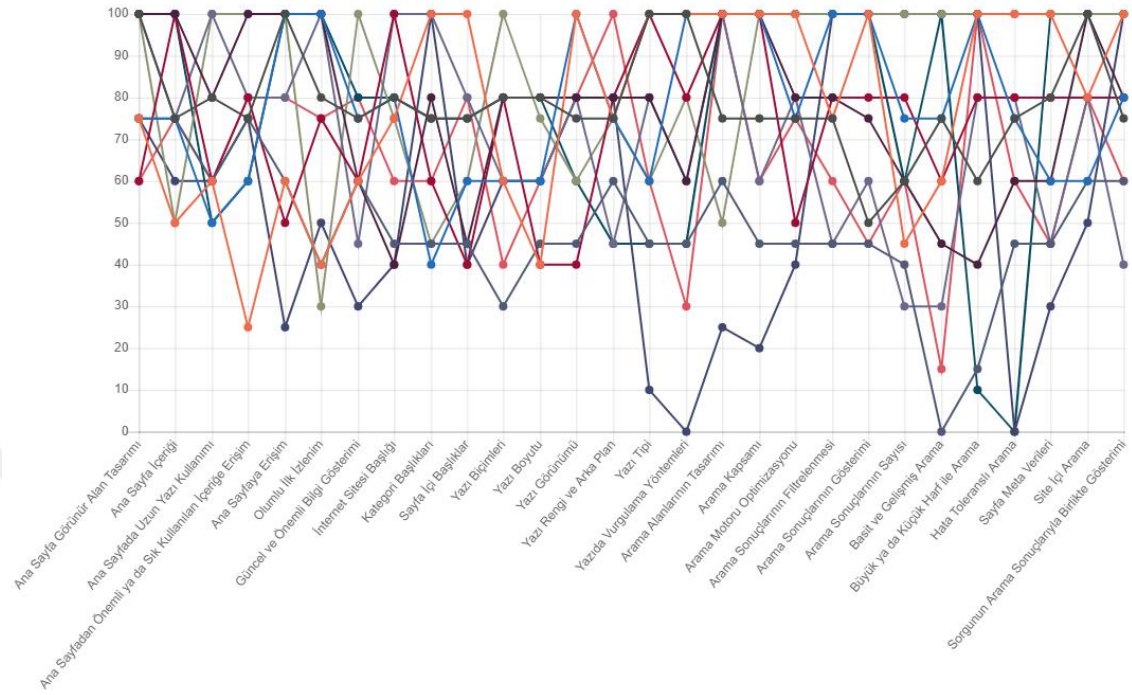
	Başlıklar Kategorisi / Değerlendirme	
İlke	**Klasik**	**Bulanık**
Yazı Biçimleri	66,36	73,64
Yazı Boyutu	61,82	76,69
Yazı Görünümü	72,73	80,48
Yazı Rengi ve Arka Plan	71,82	78,86
Yazı Tipi	67,27	74,08
Yazıda Vurgulama Yöntemleri	63,64	73,86



Şekil 5.5 Arama Kategorisine Ait İlkeler için Klasik ve Bulanık Değerlendirme Radar Grafiği

Çizelge 5.5 Arama Kategorisi, Klasik ve Bulanık Değerlendirme Sonuçları

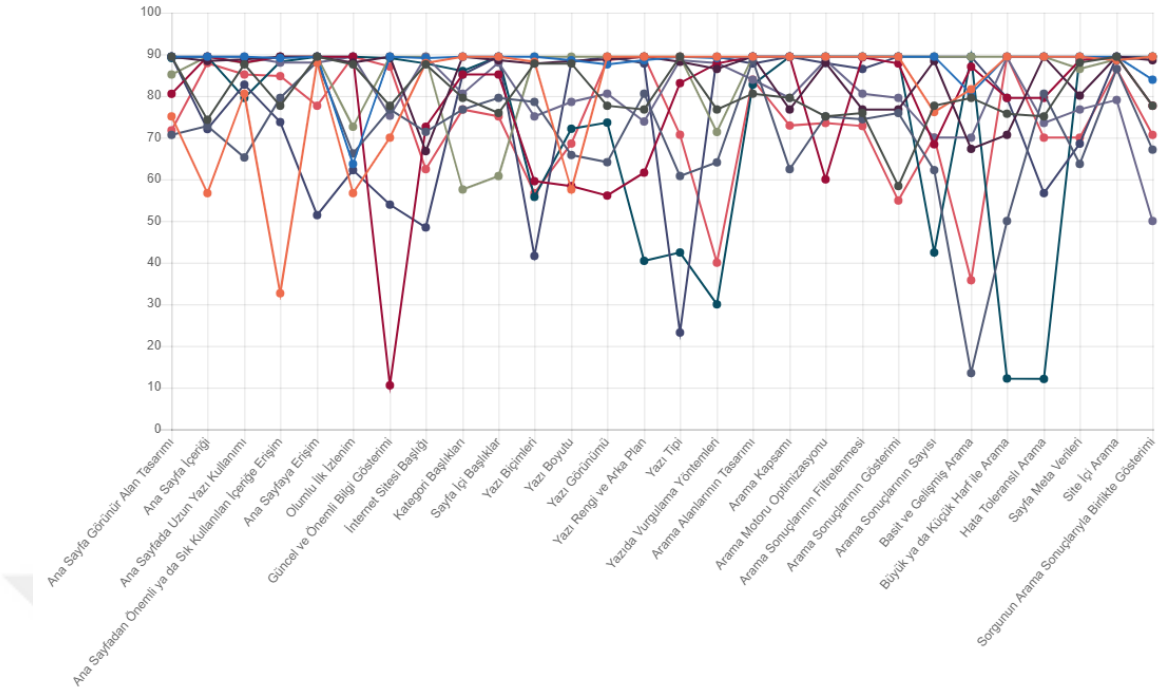
	Başlıklar Kategorisi / Değerlendirme	
İlke	**Klasik**	**Bulanık**
Arama Alanlarının Tasarımı	82,73	86,79
Arama Kapsamı	78,18	82,53
Arama Motoru Optimizasyonu	74,55	82,36
Arama Sonuçlarının Filtrelenmesi	78,18	83,09
Arama Sonuçlarının Gösterimi	77,73	80,05
Arama Sonuçlarının Sayısı	64,55	74,87
Basit ve Gelişmiş Arama	60,00	71,25
Büyük ya da Küçük Harf ile Arama	71,36	74,07
Hata Toleranslı Arama	61,36	73,21
Sayfa Meta Verileri	67,73	80,96
Site İçi Arama	80,91	87,74
Sorgunun Arama Sonuçlarıyla Birlikte Gösterimi	79,55	79,40



Şekil 5.6 Klasik Uzman Değerlendirmesi Grafiği

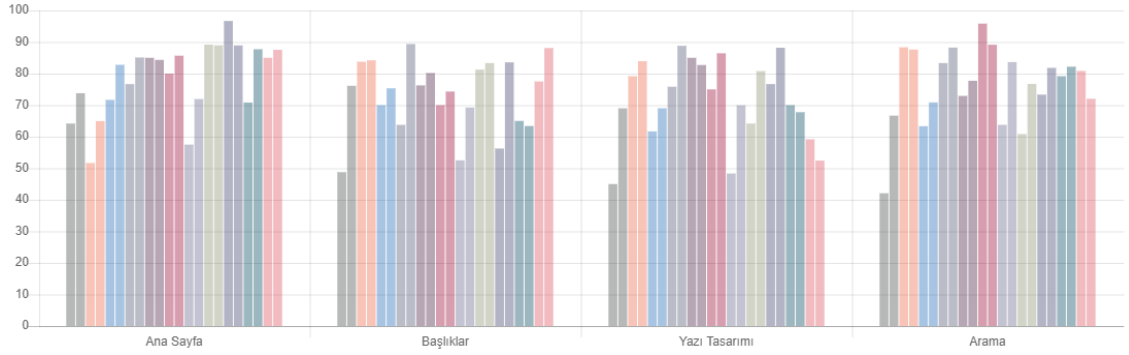
Yukarıda verilen grafik, uzman kullanıcıların ilk proje için yaptıkları klasik değerlendirmenin ilke düzeyinde sonuçlarını göstermektedir. Grafikten görüldüğü üzere değerlendirme sonuçları geniş bir aralığa dağılmış ve tutarlılık oranı düşük görünmektedir.

Bu çalışmada çizgi grafiklerin kullanılmasının nedeni; uzman kullanıcılar tarafından değerlendirilen 28 ilkenin bir bütün olarak gösterilerek, takibinin kolayca yapılabilmesi ve sapmaların daha kolay tespit edilebilmesi amacıyladır. Grafikte yer alan her bir renk farklı bir uzmanı temsil etmektedir.



Şekil 5.7 Bulanık Uzman Değerlendirmesi Grafiği

Yukarıda yer alan grafik, uzman kullanıcıların ilk proje için gerçekleştirdikleri bulanık değerlendirmeye ait ilke düzeyinde sonuçları göstermektedir. Kullanıcılar tarafından yapılan puanlamaların klasik değerlendirmeye göre dikey düzlemde daha dar bir aralıkta ve daha tutarlı olduğu görülebilmektedir. Aykırı değerler, kullanıcıların KAMİS Rehber ilkelerini farklı yorumlamalarından kaynaklanmaktadır.



Şekil 5.8 Kategori Bazında Uzman Kullanıcıların Klasik ve Bulanık Değerlendirme Sonuçlarını Gösteren Grafik

Bu grafikte ise ilk proje için uzman kullanıcıların kategori bazında klasik ve bulanık değerlendirmeler için yapmış oldukları puanlamaların ortalamaları görülmektedir. Her kullanıcı için klasik ve bulanık değerlendirme sonuçları aynı renkte ve yan yana verilmiştir. Grafik incelendiğinde kullanıcıların genel olarak her iki değerlendirme için yaptıkları puanlamalar arasındaki farklarda tutarlılık olduğu görülebiliyor. Bazı kullanıcılarda bu fark birbirine çok yakın iken bazı kullanıcılarda ise farkın daha büyük olduğu görülüyor. Bu durum, kullanıcıların her iki değerlendirme arayüzünü algılamalarındaki farklılık olarak düşünülebilir. Değerlendirme sonuçlarındaki farklılığın büyük olduğu kullanıcıların sezgisel değerlendirme yönteminden daha fazla faydalandığı söylenebilir.

KAMİS Rehber içerisinde toplam 17 kategori olmasına rağmen, bunlardan sadece 4 tanesi kullanıcıların sıkılmadan rahatça değerlendirebilmesi için seçilmiştir. KDS aracının aynı zamanda kullanılabilirlik konusunda çok fazla uzmanlığı olmayan kişileri de hedeflemesinden dolayı, ilkeler mümkün olduğunca açık ve net bir şekilde yazılarak detaylandırılmıştır. Dolayısıyla burada seçilen 4 kategori içerisinde toplam 28 ilke ve bu ilkelerle ilişkili olarak toplam 190 yönerge kullanıcılar tarafından değerlendirilmiştir. Tüm kategoriler üzerinden bakıldığında, seçilen kategoriler az görünmekle birlikte, bu değerlendirmenin yönerge seviyesinde yapıldığı düşünülürse, yeteri kadar yoğun içeriğe sahip olduğu görülebilir.

5.2 Likert Tipi Ölçek, Sürgü Ölçeği ve SKÖ Anketi Sonuçlarının Karşılaştırılması

İlk proje kapsamında toplam 11 uzman kullanıcıdan klasik ve bulanık yöntemlerle değerlendirme yapmaları istenmişti. Bu projede aynı zamanda uzman kullanıcılar da dâhil olmak üzere toplam 30 kullanıcının SKÖ anketi değerlendirmesini yapması sağlanmıştır. Kullanıcılar tarafından yapılan değerlendirmeler sonrasında, klasik / bulanık değerlendirme ve SKÖ anketi sonuçları KDS aracı üzerinden rapor sayfalarında görülebilmektedir.

SKÖ anketi 1990'lı yıllardan itibaren referans alınan ve yaygın kullanıma sahip, güvenilir bir kullanılabilirlik ölçme yöntemidir. Bu bölümde uzman kullanıcıların yapmış olduğu klasik ve bulanık değerlendirmelerden elde edilen ortalama puanlar ile SKÖ anketi puanı karşılaştırılacaktır.

Klasik değerlendirme yönteminde her ilke için önem derecesi ağırlıklandırma olarak dikkate alındığı için tüm kullanıcı değerlendirmelerinin aritmetik ortalaması alınarak nihai puanı hesaplanmıştır. Bulanık değerlendirme yönteminde ise her ilkenin puanı hesaplanırken bulanık çıkarım yöntemi kullanıldığı için ortalama puanında sadece aritmetik ortalaması alınmıştır. SKÖ anketinde her kullanıcı değerlendirmesinin sonucu, yöntemle birlikte sunulan tekniğe göre hesaplanmıştır. Tüm SKÖ anket değerlendirmelerinin ortalaması için uzman değerlendirmelerinde olduğu gibi aritmetik ortalaması alınmıştır. Değerlendirmeler sonucunda elde edilen grafik ve çizelge aşağıda görülebilir;



Şekil 5.9 Klasik / Bulanık Değerlendirme ve SKÖ Anketi Sonuçlarını Gösteren Grafik

Çizelge 5.6 Klasik / Bulanık Değerlendirme ve SKÖ Anketi Sonuçları

Yöntem	**Puan**
Uzman Değerlendirmesi (Klasik)	70,84
Uzman Değerlendirmesi (Bulanık)	**79,19**
SKÖ	**85**

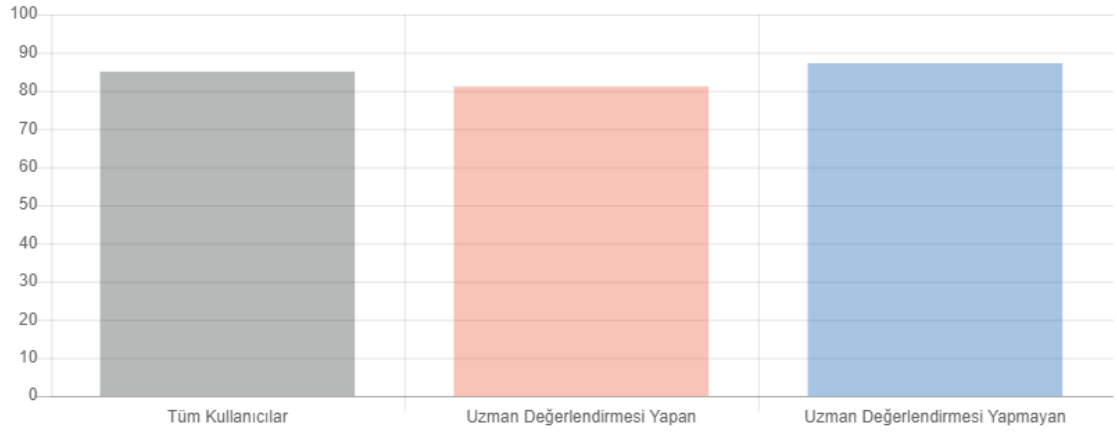
Burada yapılan deneysel çalışmanın amaçlarından birisi de bulanık yöntem ile yapılan değerlendirme sonucunun SKÖ anketi sonucuna yakınsanmış olması idi. Yukarıda yer alan grafik ve çizelgedeki veriler incelendiğinde klasik yöntemle yapılan değerlendirme sonucunda ortaya çıkan puanın 70,84 olarak tüm yöntemlerin en düşüğü olduğu görülüyor. SKÖ anketinde ise tüm kullanıcıların puanlama ortalaması 85 ile tüm değerlendirmelerin en yükseği olduğu görülüyor.

Klasik yöntemde kullanıcılara önem ve uygunluk parametreleri için sunulan kabaca 5 farklı tercihten birisini seçerek puanlama yapmaları istenmiştir. Örneğin, önem derecesi için 3 ile 4 arasında karar veremeyen bir kullanıcı, bu yönteme göre ikisinden birini tercih etmek durumunda kalmaktadır. Bu durum sonucu ciddi oranda etkileyebilmektedir.

Bulanık değerlendirme yönteminde kullanıcılara daha özgür değerlendirme yapabilmeleri için sürgü ölçeği sunulmuştu. Bu yöntemle kullanıcılar puanlamalarını sezgisel olarak kaydırarak 0 ile 100 arasında değişen bir değere denk gelecek şekilde yapabildiler. Bulanık değerlendirmeler sonucunda 79,19 puan elde edilerek klasik değerlendirmeye göre yaklaşık %10 seviyesinde daha yüksek ve SKÖ anketi sonucuna (85) yakınsanmış bir puan elde edildi.

5.3 Uzman ve Son Kullanıcı SKÖ Anketi Değerlendirmeleri

İlk projede detaylı değerlendirmeye tabi tutulan www.turkiye.gov.tr sitesinde, SKÖ anketi değerlendirmesi toplamda 30 uzman kullanıcı ve son kullanıcı tarafından gerçekleştirildi. Bu kullanıcılardan 19'u sadece SKÖ anketi üzerinden internet sitesini değerlendirdi. 11 uzman kullanıcı ise aynı zamanda detaylı ilke seti eşliğinde, daha teknik ve konuya hâkim olarak değerlendirmelerini yapmıştır. Çalışma sonucunda elde edilen grafik ve çizelge aşağıda görülebilir;



Şekil 5.10 Uzman ve Son Kullanıcı SKÖ Anketi Sonuçlarını Gösteren Grafik

Çizelge 5.7 Uzman ve Son Kullanıcı SKÖ Anketi Sonuçları

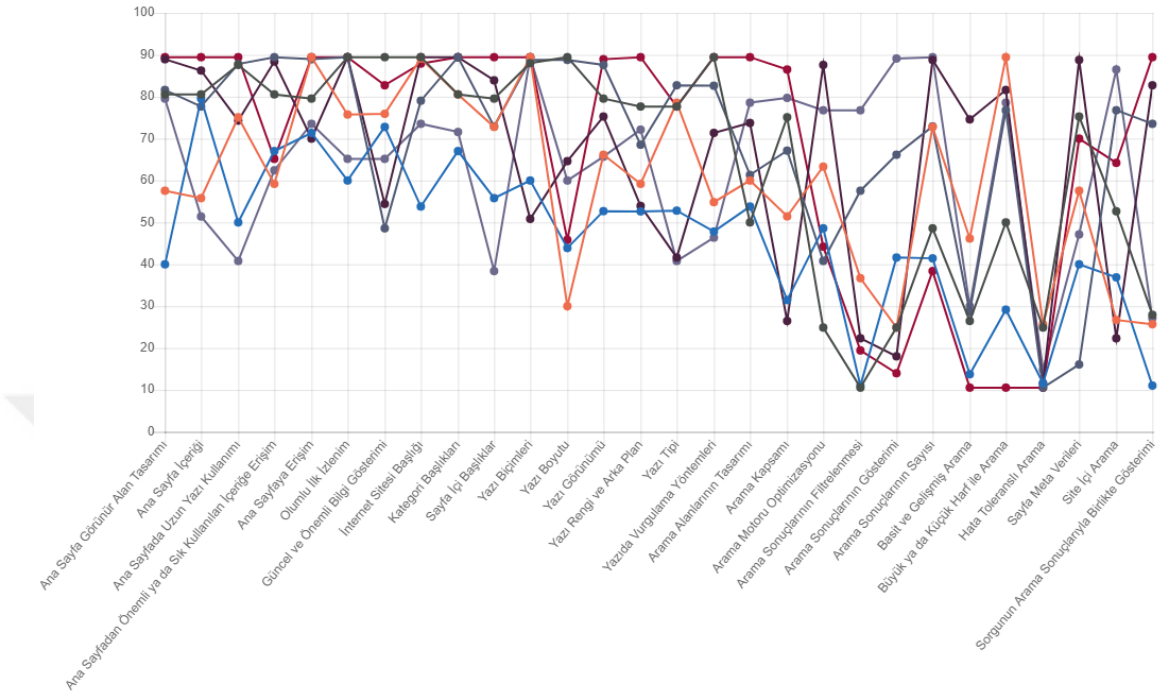
Değerlendiren	**Puan**
Tüm Kullanıcılar	85
Uzman Değerlendirmesi Yapan	**81,14**
Uzman Değerlendirmesi Yapmayan	87,24

Yukarıdaki çizelgeden görüldüğü üzere uzman değerlendirmesi yapan 11 kullanıcının SKÖ anketi değerlendirme sonucu 81,14 olarak çıkmıştır. Bu değerin önceki bölümde verdiğimiz bulanık değerlendirme sonucuna (79,19) çok yakın çıktığı görülüyor. KDS aracında tanımlı olan uzman kullanıcılar; klasik / bulanık değerlendirme ve SKÖ anket değerlendirmesi olmak üzere toplam 3 farklı değerlendirme yapmışlardır. Bunlardan bulanık değerlendirme ve SKÖ anketi değerlendirme sonuçları birbirine çok yakın çıkmış ve SKÖ anketi sonucu bulanık değerlendirme yöntemini desteklemiştir.

5.4 Tüm Projeler için Sürgü-Bulanık Değerlendirme Sonuçları

İlk projeyi toplamda 11 uzman kullanıcı klasik / bulanık değerlendirme ve SKÖ anket değerlendirmesi olmak üzere 3 farklı şekilde detaylı olarak değerlendirmişlerdi. İkinci ve üçüncü projelerde ise toplam 7 uzman kullanıcı sadece bulanık değerlendirme yapmışlardır. Bu projelerde kullanıcıların yaptığı bulanık değerlendirmenin ilk projede

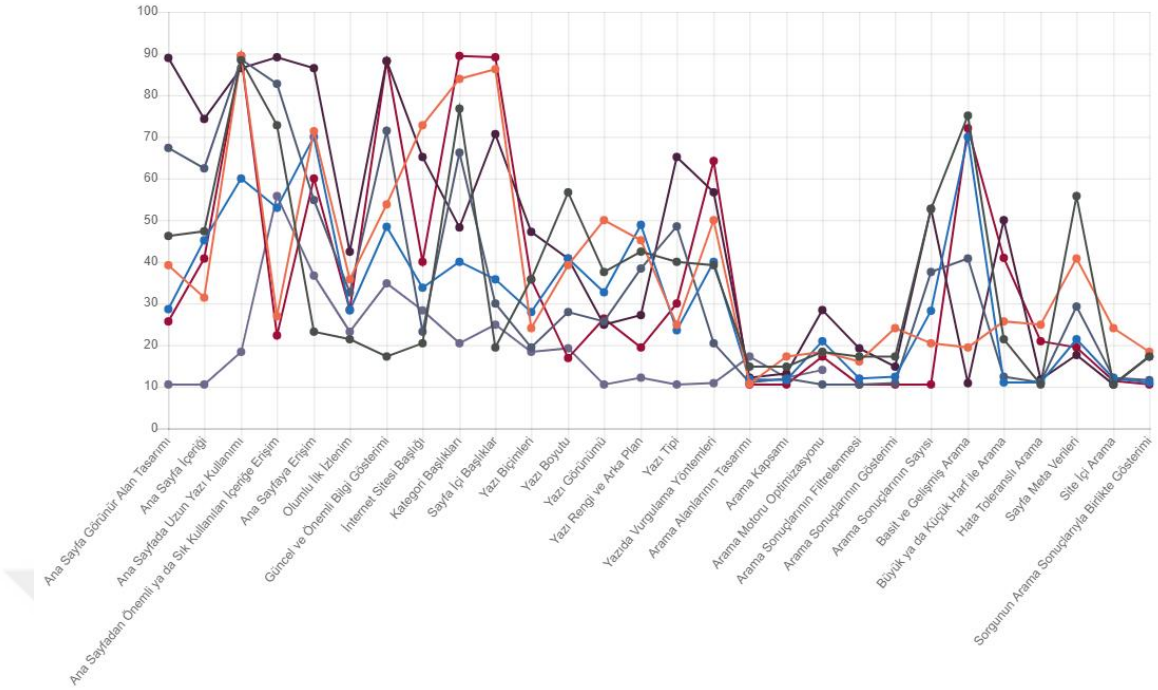
olduğu gibi daha tutarlı olması beklendi. Aşağıda yer alan grafiklerde ve çizelgede değerlendirmelere ait detaylı bilgiler verilmiştir.



Şekil 5.11 www.ssb.gov.tr internet sitesi ilke bazında bulanık uzman değerlendirmesi grafiği

www.ssb.gov.tr sitesi için yapılan değerlendirmede her kullanıcının ilke bazında verdiği puanlar çizgilerle işaretlenerek gösterilmiştir. Yukarıda yer alan grafiğe göre kullanıcıların verdiği puanlar genel olarak birbirine yakın görünmektedir. Arama kategorisine ait ilkelerde (sağ bölüm) yapılan puanlamaların daha dağınık olduğu görülüyor. Bu projede uzman değerlendirmesi yapan kullanıcılarla görüşülerek grafiğin sağ bölümünde yer alan dağınık yapının kök neden analizi yapılmıştır.

Özellikle dağınık puanlamaya sahip bazı ilkelerde, hem ilkenin ortaya koyduğu ifadenin biraz genel olduğu, hem de internet sitesinin farklı yorumlamalara elverişli olduğu anlaşılmıştır. Bu durum, KAMİS Rehber'de yer alan bazı ilkelerin daha net ifadelerle güncellenmesi gerekliliğini de ortaya çıkarmıştır.



Şekil 5.12 www.uyap.gov.tr internet sitesi ilke bazında bulanık uzman değerlendirmesi grafiği

Son projede www.uyap.gov.tr sitesi değerlendirilmiştir. Diğer internet sitelerine göre bu sitenin daha eski tarz ve farklı bir tasarıma sahip olması, bu çalışmada dikkate alınmasında etkili olmuştur. Bu sitede genel olarak tutarlı puanlamaların yapıldığı söylenebilir. Aykırı değerlere sahip ilkelerde bir önceki sitede olduğu gibi farklı yorumlamaların etkili olduğu anlaşılmıştır.

Çizelge 5.8 Tüm projeler için bulanık uzman değerlendirmesi sonuçları

	Ana Sayfa	**Başlıklar**	**Yazı Tasarımı**	**Arama**
www.turkiye.gov.tr	81,97	78,83	76,27	79,69
www.ssb.gov.tr	75,63	75,43	69,09	48,70
www.uyap.gov.tr	51,36	52,40	33,71	21,03

Tabloda her projede değerlendirilen internet sitesinin kategori bazında aldığı ortalama puanlar verilmiştir. Burada yer alan puanlar, internet sitelerinin kullanılabilirlik ve tasarım açısından ne kadar farklı olduğunu göstermektedir. Tasarım her ne kadar

kullanılabilirlik ile doğrudan ilişkili olmasa da, kullanıcıların bir internet sitesinin kalitesi ve kullanılabilirliği algısında önemli oranda etkili olmaktadır. Her iki disiplininde insan odaklı olması ve insanın kalite algısını etkilemesi, bilinçli veya bilinçaltı düzeyde insanlar tarafından değerlendiriliyor olması, birbiri ile yakın ilişkili olarak değerlendirilmesinde etkili olmaktadır.

Genel kabul görmüş kurallara ve ilkelere dikkat edilmeden, kişisel tasarım zevkine göre oluşturulmuş internet siteleri, insanlar tarafından olumsuz algılanarak başarısız bulunabilmektedir. Her insan farklı zevklere ve değerlere sahip olduğu için internet sitelerinin tasarımında, hedef kitlenin genel özellikleri göz önünde bulundurulmalı ve kullanılabilirlik konusunda standartlaşmış kural ve kaidelere dikkat edilmelidir.

Çizelge 5.9 Değerlendirilen internet sitelerinin ilkelere göre standart sapma ortalaması

İnternet Sitesi	**Standart Sapmaların Ortalaması (ilke)**
www.turkiye.gov.tr (Likert)	22,28
www.turkiye.gov.tr (Sürgü-Bulanık)	12,88
www.ssb.gov.tr (Sürgü-Bulanık)	17,21
www.uyap.gov.tr (Sürgü-Bulanık)	14,19

Yukarıda yer alan çizelgede uzman kullanıcılar tarafından değerlendirmesi yapılan 3 farklı internet sitesine ait standart sapma ortalamaları görülmektedir. Burada yer alan değerler 28 ilkenin her biri için hesaplanan standart sapma değerlerinin ortalamasıdır.

En fazla standart sapmanın 22,28 değeriyle Likert tipi ölçekte yapılan değerlendirme olduğu görülmektedir. Sürgü-bulanık yöntemine göre yapılan değerlendirmelerin tamamının bu değerin altında olduğu görülmektedir. Bu rakamlara göre sürgü-bulanık yönteminde daha tutarlı değerlendirmeler yapıldığı söylenebilir.

5.5 Likert Tipi Ölçek ve Sürgü Ölçeği için Korelasyon ve Wilcoxon Testi

11 kullanıcı örnekleminde yapılan bu çalışma ile öncelikle klasik yöntem olan Likert tipi ölçeğe göre kullanıcılardan www.turkiye.gov.tr sitesini 28 kullanılabilirlik ilkesine göre değerlendirmesi istenmiş, daha sonra ise aynı örneklem ile sürgü ölçeğine göre değerlendirme yapmaları istenmiştir. Her iki yöntemde yapılan değerlendirmeler sonucunda ortalama puanlar aşağıdaki şekilde çıkmıştır;

Çizelge 5.10 Likert tipi ölçekte ve sürgü ölçeğinde elde edilen ortalama puanlar

Yöntem	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	K11
Likert	78,57	77,86	77,50	48,39	76,61	73,39	70,54	84,29	57,50	76,61	65,71
Sürgü	80,56	81,48	87,84	70,04	84,95	77,59	81,16	85,69	76,14	73,43	73,66

K = Kullanıcı

Yukarıdaki çizelgede yer alan ortalama puan değerlerinin anlamlı bir farklılık ortaya koyduğunu ve sürgü ölçeğinin bu sonuçta etkili olduğunu göstermek amacıyla ilişkili iki örneklem testlerinden Wilcoxon işaretli sıralar testi uygulanmıştır. Test için aşağıdaki hipotezler kullanılmıştır;

H_0: Likert tipi ölçekle ve sürgü ölçeğiyle yapılan değerlendirmeler arasında anlamlı bir farklılık yoktur.

H_1: Sürgü ölçeği ile yapılan değerlendirme, sonucu etkilemiştir ve sonuçlar arasında anlamlı bir farklılık vardır.

α = 0,05

Çizelge 5.11 Wilcoxon test istatistiği

Test İstatistikleri	Sürgü - Likert
Z	-2,667
p	0,008
r	0,74

Z = Test istatistiği

p = iki uçlu önem değeri

r = korelasyon

Çizelge 5.10'da yer alan değerler üzerinden toplam 11 örneklem ile yapılan Wilcoxon testinde pozitif işaretli sayıların 10, negatif işaretli sayıların ise 1 olduğu görülmüştür. Yani 10 kişi sürgü ölçeği ile yaptığı değerlendirmede daha yüksek puanlar vermiştir. P değeri 0,008 olarak hesaplanmıştır. %95 güven düzeyinde, $p < 0,05$ olduğu için H_1 hipotezi kabul edilir. Sürgü ölçeği ile yapılan değerlendirmede anlamlı bir fark oluşmuştur. Aynı zamanda her iki yönteme göre verilen ortalama puanlar üzerinden korelasyon hesaplandığında 0,74 olduğu görülmektedir. Bu değerde her iki değerlendirmenin ilişki katsayısının yüksek olduğunu göstermektedir.

5.6 Son Puan Analizi

Son puan, her iki yönteme göre kullanıcılar tarafından girilen "Önem" ve "Uygunluk" parametreleri üzerinden hesaplanan skoru ifade etmektedir. İlk proje kapsamında toplam 11 kullanıcının 28 kullanılabilirlik ilkesini değerlendirmesi ile elde edilen verilerden, ortalama puanların, standart sapmaların ve korelasyon katsayısının hesaplanması ile bir analiz gerçekleştirilmiştir. Bu bölümde elde edilen veriler çizelge halinde paylaşılmış ve değerlendirmeler yapılmıştır.

Likert tipi ölçeğe göre klasik değerlendirme yapan kullanıcıların her ilke için verdikleri puanlarda hesaplanan standart sapma oranı, sürgü ölçeğinin ve bulanık değerlendirme yönteminin kullanıldığı ikinci değerlendirmede verilen puanlara göre çok daha yüksek oranlarda çıktığı görülmüştür. Likert tipi ölçek yönteminde toplam 28 ilkenin standart sapmalarının ortalaması 22,28 olarak çıkmış, sürgü-bulanık yönteminin kullanıldığı durumda ise 12,88 olduğu görülmüştür. Buradaki rakamlara göre ikinci yöntemde daha tutarlı puanlar elde edildiği görülmektedir.

Yapılan analize dair detaylı bilgi aşağıdaki çizelgede verilmiştir;

Çizelge 5.12 Likert tipi ölçek ve Sürgü-Bulanık yöntemlerine göre Son Puan analizi

İlke	**Likert**		**Sürgü-Bulanık**		**r**
	Ort.	Std. sapma	Ort.	Std. sapma	
Ana Sayfa Görünür Alan Tasarımı	83,64	15,82	83,59	7,34	0,66
Ana Sayfa İçeriği	75,91	17,43	81,76	10,60	0,56
Ana Sayfada Uzun Yazı Kullanımı	69,09	17,30	84,04	6,83	0,41
Ana Sayfadan Önemli ya da Sık Kullanılan İç. E.	73,64	19,67	80,18	15,93	0,78
Ana Sayfaya Erişim	77,73	24,62	84,54	10,98	0,68
Olumlu İlk İzlenim	71,82	26,14	77,71	12,76	0,63
Güncel ve Önemli Bilgi Gösterimi	65,91	18,07	73,52	22,52	0,47
İnternet Sitesi Başlığı	70,45	20,61	77,55	13,42	0,70
Kategori Başlıkları	70,91	21,72	81,44	8,84	0,43
Sayfa İçi Başlıklar	63,64	18,96	82,81	8,76	**0,00**
Yazı Biçimleri	66,36	19,20	73,64	16,43	**0,22**
Yazı Boyutu	61,82	14,81	76,69	12,08	0,75
Yazı Görünümü	72,73	18,75	80,48	11,05	0,81
Yazı Rengi ve Arka Plan	71,82	15,41	78,86	14,86	0,62
Yazı Tipi	67,27	26,91	74,08	21,76	0,83
Yazıda Vurgulama Yöntemleri	63,64	30,75	73,86	20,11	0,36
Arama Alanlarının Tasarımı	82,73	25,35	86,79	3,18	**-0,14**
Arama Kapsamı	78,18	26,99	82,53	8,70	0,44
Arama Motoru Optimizasyonu	74,55	20,61	82,36	9,46	0,55
Arama Sonuçlarının Filtrelenmesi	78,18	20,14	83,09	6,67	0,71
Arama Sonuçlarının Gösterimi	77,73	22,80	80,05	12,17	0,86
Arama Sonuçlarının Sayısı	64,55	21,58	74,87	13,92	0,49
Basit ve Gelişmiş Arama	60,00	33,03	71,25	23,57	0,89
Büyük ya da Küçük Harf ile Arama	71,36	33,45	74,07	22,68	0,88
Hata Toleranslı Arama	61,36	32,83	73,21	21,68	0,80
Sayfa Meta Verileri	67,73	24,34	80,96	9,27	0,83
Site İçi Arama	80,91	17,30	87,74	2,95	**0,08**
Sorgunun Arama Sonuçlarıyla Birlikte Gösterimi	79,55	19,12	79,40	12,08	0,95
Ortalama	**71,54**	**22,28**	**79,32**	**12,88**	**0,58**

Yukarıda yer verilen çizelgeye göre “Sayfa İçi Başlıklar”, “Yazı Biçimleri”, “Yazıda Vurgulama Yöntemleri”, “Arama Alanlarının Tasarımı” ve “Sayfa İçi Arama” gibi ilkelerden

elde edilen korelasyon katsayı değerinin diğerlerine oranla daha düşük olduğu görülmektedir. Özellikle "Arama Alanlarının Tasarımı" ilkesi için negatif korelasyon katsayısının olması, bu ilkede kişilerin düşükte olsa tersi istikamette değerlendirme yaptığını göstermektedir. Arama ile ilgili olarak değerlendirilen diğer ilkelerin pozitif korelasyon katsayı değerine sahip olması, bu ilkede aykırı değer çıktığı şeklinde yorumlanabilir.

"Sayfa İçi Başlıklar", "Yazı Biçimleri" ve "Site İçi Arama" ilkelerinde düşük korelasyon katsayı değerinin çıktığı görülmektedir. Bunlardan ilki için kullanıcıların dikkate aldığı yönergeler incelendiğinde, diğer ilkelere nispeten daha soyut şekilde ifade edilmiş yönergelere sahip olduğu (Örn. "Kullanıcılar için faydalı olduğu mesajını veren başlıklar seçilmelidir.") görülmüştür. Düşük korelasyona sahip ikinci ilkede ise yönergelerdeki ifadelerin değerlendirilen internet sitesinde tam karşılığının olmamasının etkili olduğu düşünülebilir. Son ilkede ise çok fazla yönergenin olması, kullanıcıların vereceği nihai puandaki sapma oranının fazla olmasında etkili olmuştur denebilir.

Aynı zamanda, Likert tipi ölçekte yapılan değerlendirmelerin daha düşük ortalamaya ve daha yüksek standart sapma oranlarına sahip olduğu görülmektedir. Bu durum, sürgü-bulanık yönteminin daha tutarlı sonuçlar ürettiğini göstermektedir. Likert tipi ölçekten elde edilen puana (71,54) göre daha yüksek bir puan (79,32) elde edilmiş ve bu puanın SKÖ anketindeki uzman değerlendirmeleri ortalamasına (81,14) çok daha yakın bir puan olduğu görülmüştür.

5.7 Önem Parametresi Analizi

Her bir ilke için kullanıcıların seçmiş olduğu parametrelerden birisi olan "Önem" için istatistiki veriler üzerinden analiz yapılmıştır. Genel olarak korelasyon katsayısının yüksek olduğu ve her iki yöntemde bu parametre için kullanıcıların seçmiş olduğu değerin yakın ilişkili olduğu görülmüştür.

Analiz sonuçlarına ait çizelge aşağıda paylaşılmıştır;

Çizelge 5.13 Likert tipi ölçek ve Sürgü-Bulanık yöntemlerine göre Önem parametresi analizi

İlke	**Likert**		**Sürgü-Bulanık**		**r**
	Ort.	Std. sapma	Ort.	Std. sapma	
Ana Sayfa Görünür Alan Tasarımı	4,82	0,39	88,45	11,15	**0,36**
Ana Sayfa İçeriği	4,91	0,29	92,73	8,58	0,84
Ana Sayfada Uzun Yazı Kullanımı	4,09	0,79	79,00	13,26	0,76
Ana Sayfadan Önemli ya da Sık Kullanılan İç. E.	4,45	0,66	90,64	7,52	0,48
Ana Sayfaya Erişim	4,64	0,48	90,00	12,31	0,88
Olumlu İlk İzlenim	4,55	0,66	80,91	12,54	**0,21**
Güncel ve Önemli Bilgi Gösterimi	4,00	0,74	76,82	20,80	0,85
İnternet Sitesi Başlığı	4,00	0,95	74,45	17,91	0,45
Kategori Başlıkları	4,27	0,75	81,27	17,28	0,77
Sayfa İçi Başlıklar	4,09	0,67	81,64	14,42	0,56
Yazı Biçimleri	3,55	0,89	65,45	20,57	0,61
Yazı Boyutu	3,82	0,57	71,73	13,46	0,73
Yazı Görünümü	4,18	0,57	79,91	15,69	0,36
Yazı Rengi ve Arka Plan	4,18	0,83	79,82	22,10	0,79
Yazı Tipi	3,82	0,94	69,36	30,22	0,85
Yazıda Vurgulama Yöntemleri	3,82	0,94	68,18	28,09	0,77
Arama Alanlarının Tasarımı	4,91	0,29	89,55	8,76	0,52
Arama Kapsamı	4,55	0,66	87,82	16,64	0,92
Arama Motoru Optimizasyonu	4,55	0,66	89,91	12,79	0,85
Arama Sonuçlarının Filtrelenmesi	4,27	0,86	84,45	16,94	0,97
Arama Sonuçlarının Gösterimi	4,45	0,78	87,45	14,62	0,90
Arama Sonuçlarının Sayısı	3,55	1,08	69,09	24,42	0,82
Basit ve Gelişmiş Arama	4,00	1,04	79,36	20,29	0,92
Büyük ya da Küçük Harf ile Arama	3,73	1,54	70,27	30,24	0,92
Hata Toleranslı Arama	4,27	0,86	81,82	18,81	0,65
Sayfa Meta Verileri	3,82	0,94	75,36	19,78	0,93
Site İçi Arama	4,45	0,50	87,45	10,34	0,47
Sorgunun Arama Sonuçlarıyla Birlikte Gösterimi	4,09	1,00	73,18	25,32	0,98
Ortalama	**4,21**	**0,76**	**80,22**	**17,32**	**0,72**

Yukarıdaki çizelgeden korelasyon katsayıları incelendiğinde en düşük değere sahip iki ilkenin “Ana Sayfa Görünür Alan Tasarımı” ve “Olumlu İlk İzlenim” olduğu

görülmektedir. Bu ilkelerin diğerlerine oranla daha fazla yoruma açık olduğu ve korelasyon katsayılarının düşük olmasında bu durumun etkili olduğu söylenebilir.

5.8 Uygunluk Parametresi Analizi

Uygunluk parametresinden elde edilen veriler üzerinden yapılan analiz sonuçlarına göre her iki yöntemde kullanıcılar tarafından seçilen değerlerin korelasyon katsayı ortalamasının düşük olduğu görülmüştür. Bu parametre için elde edilen değerlerin son puan ve önem derecesi parametresinden elde edilen değerlerden daha düşük olduğu ve kullanıcıların Likert tipi ölçek yöntemine göre seçtikleri değerleri, sürgü-bulanık yönteminde orantılı olarak değerlendirmedikleri ve bundan dolayı daha farklı bir dağılımın ortaya çıktığı görüldü.

Uygunluk parametresine ait analiz sonuçlarını gösteren çizelge aşağıdadır;

Çizelge 5.14 Likert / Sürgü-Bulanık yöntemlerinde Uygunluk parametresi analizi

İlke	**Likert**		**Sürgü-Bulanık**		**r**
	Ort.	Std. sapma	Ort.	Std. sapma	
Ana Sayfa Görünür Alan Tasarımı	3,45	0,50	88,18	9,96	0,53
Ana Sayfa İçeriği	3,09	0,67	84,91	12,68	0,42
Ana Sayfada Uzun Yazı Kullanımı	3,45	0,78	90,45	6,64	0,36
Ana Sayfadan Önemli ya da Sık Kullanılan İç. E.	3,36	0,88	85,18	18,67	0,96
Ana Sayfaya Erişim	3,36	0,98	91,36	13,21	0,87
Olumlu İlk İzlenim	3,09	0,90	82,00	16,37	0,71
Güncel ve Önemli Bilgi Gösterimi	3,27	0,62	80,45	26,33	0,35
İnternet Sitesi Başlığı	3,55	0,66	84,00	14,04	0,12
Kategori Başlıkları	3,27	0,62	85,73	11,03	0,18
Sayfa İçi Başlıklar	3,09	0,67	87,64	10,17	0,14
Yazı Biçimleri	3,73	0,45	84,36	17,24	-0,18
Yazı Boyutu	3,27	0,75	85,00	16,30	0,69
Yazı Görünümü	3,45	0,66	86,00	13,80	0,84
Yazı Rengi ve Arka Plan	3,45	0,50	86,18	10,91	-0,02

Çizelge 5.15 Likert / Sürgü-Bulanık yöntemlerinde Uygunluk parametresi analizi (devamı)

İlke	Likert		Sürgü-Bulanık		r
	Ort.	Std. sapma	Ort.	Std. sapma	
Yazı Tipi	3,36	0,88	85,36	17,51	0,83
Yazıda Vurgulama Yöntemleri	3,27	1,14	84,09	16,35	-0,03
Arama Alanlarının Tasarımı	3,36	0,98	93,00	6,40	-0,28
Arama Kapsamı	3,36	0,88	88,45	9,54	0,06
Arama Motoru Optimizasyonu	3,27	0,75	86,64	12,09	0,67
Arama Sonuçlarının Filtrelenmesi	3,64	0,48	88,64	7,90	0,59
Arama Sonuçlarının Gösterimi	3,45	0,66	84,91	15,70	0,89
Arama Sonuçlarının Sayısı	3,64	0,48	86,18	10,44	0,09
Basit ve Gelişmiş Arama	2,73	1,21	74,73	25,94	0,94
Büyük ya da Küçük Harf ile Arama	3,64	0,64	83,73	22,37	0,82
Hata Toleranslı Arama	3,00	1,48	82,45	25,67	0,91
Sayfa Meta Verileri	3,45	0,50	89,36	8,78	0,65
Site İçi Arama	3,64	0,64	93,55	5,47	0,16
Sorgunun Arama Sonuçlarıyla Birlikte Gösterimi	3,91	0,29	93,09	5,81	0,71
Ortalama	**3,38**	**0,74**	**86,27**	**13,83**	**0,46**

Analiz sonuçlarına göre tüm ilkelerin her iki yönteme göre korelasyon katsayı ortalaması 0,46 olarak hesaplanmıştır. Korelasyon katsayısının düşük olması, değerlendirmeyi yapan kullanıcılarla tartışılmış ve nedenleri araştırılmıştır. Bu konuyla ilgili olarak değerlendirmeleri yapan kullanıcılara yönelik bir anket düzenlenerek fikirleri alınmış ve bu farklılığın oluşmasındaki nedenler incelenmiştir. En öne çıkan fikir; "Önem" parametresinin Likert tipi ölçek yönteminde 1-5 arası rakamsal değerler olarak kullanıcılara sunulmuş olması, "Uygunluk" parametresinin ise "Uygun değil" ile "Uygun" arasında değişen metinsel ifadeler olarak sunulması olmuştur. Kullanıcılar Likert tipi ölçek yöntemindeki rakamsal değerler ile daha rahat ve net karar verebildiklerini, bundan dolayı sürgü ölçeği ile benzer değerler çıktığını, "Uygunluk" parametresinde kullanılan metinsel değerlerin algılanmasının nispeten zor olması ve zihinlerinde rakamsal karşılığını tam oluşturamamalarından dolayı sürgü ölçeğinde çok farklı değerler seçebildiklerini belirtmişlerdir.

5.9 Güvenilirlik ve İç Tutarlılık Analizi

Kullanıcıların değerlendirdiği 28 ilke için güvenilirlik ve iç tutarlılığın ölçülmesi amacıyla Cronbach's α yönteminden (Cronbach 1951) istifade edilmiştir. Bu yöntemle "Önem" ile "Uygunluk" parametreleri ve Son Puan değeri için hem Likert tipi ölçek yönteminde, hem de sürgü-bulanık yönteminde Cronbach's α değerleri hesaplanmıştır. Cronbach's α değerleri aşağıdaki çizelgeden görülebilir;

Çizelge 5.16 Likert tipi ölçek ve Sürgü-Bulanık yöntemlerine göre güvenirlik ve iç tutarlılık analizi

Değerlendirme Yöntemi	**Değer**	**Cronbach's α**
Likert	Önem	0,92
Sürgü-Bulanık		0,92
Likert	Uygunluk	0,75
Sürgü-Bulanık		0,63
Likert	Son Puan	0,84
Sürgü-Bulanık		0,78

[1] Likert tipi ölçek ve sürgü-bulanık yöntemine göre hesaplanan ortalama puanlar arasındaki korelasyon katsayısıdır.

Güvenilirlik kategorilerinin sınıflandırılmasına dair bilgiler aşağıdaki çizelgede paylaşılmıştır. Cronbach's α değerleri aşağıdaki çizelgeye göre yorumlanabilir;

Çizelge 5.17 Cronbach's α güvenilirlik kategorileri

Cronbach's α	**Güvenilirlik**
$\alpha \geq 0,9$	Mükemmel
$0,9 > \alpha \geq 0,8$	İyi
$0,8 > \alpha \geq 0,7$	Kabul edilebilir
$0,7 > \alpha \geq 0,6$	Tartışmaya açık
$0,6 > \alpha \geq 0,5$	Kötü
$0,5 > \alpha$	Güvenilmez

Çizelge 5.16'te görüldüğü üzere "Önem" parametresinin her iki yöntemde yapılan değerlendirmelere ait güvenilirliğin ve iç tutarlılığın 0,92 ile çok iyi çıktığı görülmektedir. Kullanıcılarla yapılan görüşmelerde "Önem" parametresinin zihinlerdeki karşılığının daha net olduğu ve kolayca aynı değerlendirmeyi tekrar yapabildikleri belirtilmiştir. "Uygunluk" parametresi için güvenilirlik değerleri, Likert tipi ölçek ve sürgü-bulanık yöntemine göre sırasıyla 0,75 ve 0,63 çıkmıştır. Bu değerler Çizelge 5.17'da yer alan güvenilirlik kategorilerinden incelendiğinde, sırasıyla "Kabul edilebilir" ve "Tartışmaya açık" çıktığı görülmektedir.

"Uygunluk" parametresi için 11 kullanıcının sürgü-bulanık yönteminde yapmış olduğu toplam 308 değerlendirmede, ortalamadan çok uzak olan aykırı değerlerden sadece 3 tanesi (7, 24 ve 25. ilkeler) değiştirildiğinde, sonucun kabul edilebilir seviyelere çıktığı görülmüştür.

5.10 Kullanıcı Geribildirimleri

Belirli aşamalarda uzman kullanıcılarla ikili görüşmeler yapılmış ve çalışma sonunda kısa bir anket değerlendirmeleri istenmiştir. Bu çalışmalarla, kullanıcıların KDS sistemi ile ilgili kullanım deneyimleri ve genel olarak sonuçlar üzerine fikirleri alınmıştır.

Çalışma sonundaki ankette toplam 3 soru sorulmuştur. Bu sorulardan ilk ikisi çoktan seçmeli olarak yöntemin değerlendirilmesine yönelik, son soru ise açık uçlu cevap şeklinde tasarlanmıştır. Sorular aşağıda görülebilir;

1. Hangi değerlendirme yöntemi daha eğlenceli ve ilgi çekiciydi?
2. Hangi değerlendirme yönteminde daha özgürce ve istediğiniz puanı verebildiğinizi hissettiniz?
3. Her iki yönteme (Seçim Kutusu ve Sürgü) göre "Önem" ve "Uygunluk" parametrelerini değerlendirdiniz. "Önem" her iki yöntemde tutarlı çıktı ama "Uygunluk" parametresi için iki yöntemde çok farklı puanlar verdiğiniz görüldü. Bunun nedeni ne olabilir?

Anket toplam 9 uzman kullanıcı tarafından tamamlanmıştır. İlk soruda, 6 kullanıcı sürgü ölçeği yöntemini, 3 kullanıcı ise Likert tipi ölçeği seçmiştir. İkinci soruda, 7 kullanıcı sürgü ölçeğini, 2 kullanıcı ise Likert tipi ölçeği seçmiştir. Üçüncü soru için öne çıkan fikir ise Likert tipi ölçek yönteminde "Önem" parametresi için sayısal değerlerin kullanılmış olması, "Uygunluk" parametresi içinse sözel değerlerin kullanılmış olması yönündedir. Likert tipi ölçek yönteminde seçilen sayısal değerlerin eşdeğerinin sürgü yönteminde daha kolay seçilebildiğini, ancak sözel değerlerin sürgü ölçeğinde karşılığını seçmenin daha zor olduğunu belirtmişlerdir. Bunun dışında bir diğer öne çıkan yorum ise "Önem" parametresinin, "Uygunluk" parametresine göre daha somut bir kavram olduğu ve bundan dolayı bu parametrede daha tutarlı sonuçların çıktığı yönünde olmuştur.

İnternet sitelerinin değerlendirilmesinde referans ilke seti KAMİS Rehber'den alınmıştı. Çalışma esnasında Rehber ilkelerine yönelik yorumlar da gelmiştir. Kullanıcıların bir kısmı bazı ilkelerin daha kolay anlaşılır hale getirilebileceği konusunda önerilerde bulunmuşlardır.

5.11 Deney Sonuçlarının Önceki Çalışmalar ile Karşılaştırılması

Literatür bölümünde de paylaşılan çalışmalardan iki tanesi, bu çalışma kapsamında referans alınan KAMİS Rehber'den önce hazırlanan ve selefi konumunda olan KAKİS Rehber referans alınarak yapılmıştır. Bu çalışmalardan birisi (Durmuş ve Çağıltay 2011) araştırmacı tarafından 33 kamu internet sitesinin değerlendirilmesi ile diğeri (Arsoy vd. 2013) ise 3 farklı değerlendiricinin 6 farklı internet sitesini değerlendirmesi ile gerçekleştirilmiştir. Bu tez çalışmasında ise 3 farklı kamu internet sitesi değerlendirilmiştir. Bunlardan ilki 11 farklı uzman kullanıcı ve 30 son kullanıcı tarafından değerlendirilmiş, diğer ikisi ise 9'ar uzman kullanıcı tarafından değerlendirilmiştir.

İlk çalışmada değerlendirilen internet sitelerinin %61,2 ile %88,8 arasında değişen performansları olduğu ölçülmüştür. İkinci çalışmada ise internet siteleri için %56,65 ile

%77,26 arasında değişen ortalama puanlar elde edilmiştir. Bu tez çalışmasında değerlendirilen internet siteleri için ise %39,63 ile %79,19 arasında değişen ortalama puanlar elde edilmiştir. Burada yapılan çalışmada değerlendirici sayısı daha fazla olduğu halde elde edilen ortalama puanların proje bazında daha değişken olabildiği görülmüştür. Bu sonuçta etkili olan faktörlerden bir tanesi, seçilen internet sitelerinden birinin özellikle güncel olmayan bir tasarıma sahip olması ve bundan dolayı değerlendiricilerden düşük puanlar alması olarak açıklanabilir. Diğer etken ise sistemin, değerlendiricilerin kararlarını esnek bir şekilde vermelerine imkân tanıması olarak ifade edilebilir.

Literatürde, Likert tipi ölçek ve sürgü ölçeklerinin karşılaştırılmasına yönelik yapılan çalışmalarda her iki yöntem açısından da hem olumlu hem de olumsuz sonuçlar ortaya çıkarılmıştır. KDS sistemi üzerinden yapılan değerlendirmede, bu iki yöntemden elde edilen ortalama puanlar sırası ile 70,84 ile 69,58 olmuştur. Daha önce yapılan bir araştırmada (Toepoel ve Funke 2018) ifade edildiği gibi sürgü ölçeğinin (bulanık mantık kuralları uygulanmadan) kullanıldığı değerlendirmede nispeten daha düşük bir ortalama puan elde edilmiştir. Ancak bu çalışmada Likert ölçeğindeki dezavantajlı durumlar göz önünde bulundurularak kullanıcıların daha özgür değerlendirme yapabilmeleri için sürgü ölçeği kullanılmıştır. Sürgü ölçeği yöntemi, bulanık çıkarım sistemi ile desteklenerek daha ideal sonuçların elde edilmesi sağlanmıştır. Her iki yöntem, uzman kullanıcıların SKÖ anket değerlendirmesinden elde edilen puanla karşılaştırılmış ve sürgü-bulanık yönteminden elde edilen skorun anket skoruna çok yakın bir değer çıkardığı görülmüştür.

Bu tez çalışması, farklı açılardan daha önce yapılan Likert tipi ölçek ve sürgü ölçeği karşılaştırma çalışmalarıyla kıyaslanmıştır. Karşılaştırmaya dair çizelge aşağıda görülebilir.

Çizelge 5.18 Likert / Sürgü Ölçeği yöntemlerinde karşılaştırma çalışmaları

	Tez Çalışması	**Cape 2009**	**Roster vd. 2015**	**Toepel ve Funke 2018**
Kullanılan veri seti büyüklüğü	28 ilke	4 soru	29 soru	26 soru
Uygulama alanı (neyi ölçtüğü)	Kullanılabilirlik	Kişisel Değerler	Yönetici Tepkileri	Farklı Ölçek ve Cihaz Kullanımı
Online [2]	✓	✓	✓	✓
Kullanıcı profili	Uzman / Son Kullanıcı	Son Kullanıcı	Son Kullanıcı	Son Kullanıcı
Sürgü ölçeği - düşük ortalama puan	✓	Fark Yok [4]	✓	✓
Sürgü ölçeği - düşük cevaplama oranı	GD [3]	GD	Fark Yok	✓
Sürgü ölçeği – eğlenceli mi?	✓	✓	Fark Yok	GD

[1] Çalışmanın aynı kullanıcılarla tekrarlandığını ifade eder.

[2] Çalışmanın webden yapıldığını ifade eder.

[3] Geçerli Değil. Çalışmada ölçülmediğini veya bu konuda çıkarım yapılamayacağını gösterir.

[4] İstatistiki olarak anlamlı fark yok.

Cape (2009) yaptığı çalışmada, sürgü ölçeklerinin geleneksel Likert tipi ölçeklere göre nispeten daha eğlenceli olduğunu ve kullanıcıların gerçek düşüncelerini sürgü ölçeklerinde daha rahat yansıtabildiklerini ifade etmiştir. Sürgü ölçeğinden elde edilen puan ortalamasının Likert tipi ölçekten istatistiki olarak anlamlı bir farklılığa sahip olmadığını göstermiştir. Çalışmada, Likert tipi ölçeğe göre değerlendirme yapan kullanıcıların, ikinci bir değerlendirme yaparak yanıtlarını güncellemelerine imkân tanınmıştır. Kullanıcıların birçoğunun ilk değerlendirme de vermiş oldukları yanıtları değiştirdikleri görülmüştür. Aynı zamanda farklı görsellikte sürgü ölçekleri kullanılarak gelen yanıtların farklılaşması incelenmiştir. Buna göre kullanıcıların yanıtlarının sadece Likert tipi ölçek veya sürgü ölçeği gibi yöntem değiştiğinde değil, aynı zamanda sürgü ölçeğinde kullanılan görsel temaya göre de farklılaştığı görülmüştür. Kullanıcı algısının

ve buna göre değerlendirmesinin öncelikli olarak yönteme göre daha sonra görsellik ve içerik gibi daha detaysal farklılıklardan etkilendiği bu gibi çalışmalarda görülebilmektedir. Bu tez çalışmasında farklılıklar üzerinde en fazla etkisi olan durumlardan birisi olarak yöntem değiştirilmiş ve sonuçlar karşılaştırmalı olarak incelenmiştir.

Sürgü ölçeklerinin daha eğlenceli olduğu ve veri kalitesinin Likert tipi ölçeklere eşdeğer veya daha üstün olduğu argümanı ile yapılan bir diğer çalışmada (Roster vd. 2015), her iki yöntem arasında istatistiki olarak anlamlı bir farklılığın olmadığı sonucu çıkarılmıştır. Çalışmada katılımcılara 3 başlık altında toplam 29 soru yöneltilmiş ve ortalamalar üzerinden istatistiki hesaplamalar yapılarak çıkarımlarda bulunulmuştur. Yapılan çalışmalarda katılımcılardan sürgü ölçeği anketi ile elde edilen katılım skorunun, klasik yöntem olan Likert tipi radyo butonları ile sunulan anketten elde edilen skorlara göre, özellikle 25-34 yaş aralığında daha yüksek olduğu ve bu farklılığın istatistiki olarak anlamlı olduğuna değinildiği belirtilmiştir. Sürgü ölçeği ile anketi tamamlama süresinin, klasik yönteme oranla daha uzun olduğu görülmüş ancak bunun istatistiki olarak anlamlı olmadığı belirtilmiştir. Genel olarak bu yöntem karşılaştırmasında cevaplama oranı, tamamlama süresi ve veri kalitesi gibi değişkenler açısından sonuç olarak sunulabilecek belirgin farklılığın olmadığı görülmüştür.

Toepoel ve Funke (2018) tarafından yapılan çalışmada, katılımcılara 5, 7 veya 11'li ölçeklerde, kişisel bilgisayar, tablet veya cep telefonunda anket sorularını cevaplamaları istenmiş ve ortalama skorlar ile yanıtlanmayan soru sayılarındaki değişim incelenmiştir. Kişisel bilgisayarlarda cevaplanmayan soru sayısının en az olduğu görülmüş ve mobil cihazlara göre daha düşük ortalama puan elde edildiği görülmüştür. Sürgü ölçeğinin daha düşük ortalama skorlara ve daha fazla cevaplanmayan soruya neden olduğu görülmüştür. 5'li ölçeklerin, 11'li ölçeklere göre daha fazla cevaplanmayan soruya neden olduğu ve daha olumsuz değerlendirildiği görülmüştür. Taşı-bırak tarzındaki sürgü ölçeklerinin radyo butonlarından, büyük butonlardan ve görsel analog ölçeklerden daha kötü sonuç verdiği görülmüştür. Katılımcılar 11 noktalı sürgü ölçeğinin 5 ve 7 noktalı sürgü ölçeğinden daha zor olmadığını, hatta işlerini daha da kolaylaştırdığını belirtmişlerdir. Ancak 11 noktalı ölçeğin mobil cihazlarda sorun oluşturacak derecede

daha fazla yer kapladığı belirtilmiştir. Katılımcıların anketleri değerlendirmek için mobil cihazdan daha çok masaüstü cihazları tercih ettiği görülmüştür.

Genel olarak sürgü ölçeği ve klasik Likert tipi ölçek karşılaştırma çalışmalarında, sürgü ölçeğinin daha düşük ortalama puan ortaya çıkardığı, daha uzun sürede değerlendirilebildiği ve daha düşük tamamlanma oranlarına sahip olduğu ifade edilmiştir. Bununla birlikte insanların sürgü ölçeğini daha eğlenceli bulduğu ve daha özgür değerlendirme yapmalarına imkân tanıdığı ifade edilmiştir. Elde edilen skorların, değerlendirmenin tekrar edilmesi, aynı değerlendirme yönteminin farklı görselliklerle sunulması gibi faktörlerden etkilendiği farklı çalışmalarla görülmüştür. Bu çalışmada sürgü ölçeğinin kullanımı, katılımcılara özgür değerlendirme şansı tanımasından dolayı tercih edilmiş ve bu değerlendirmeleri zenginleştirmek için elde edilen skorlar bulanık çıkarım yöntemine göre işlemden geçirilmiş, belirlenen kümeler için ara değerler bulanıklaştırılmıştır.

Bu tür çalışmalarda araştırmacıların genelde örneklem gruplar üzerinden sadece farklı yöntemleri karşılaştırmaya yönelik çalışmalar yaptığı görülmüştür. Bu tez çalışmasında ise psikolojik faktörler açısından da araştırma yapılmıştır. Özellikle sürgü ölçeğinin görünümünde bu araştırmalar dikkate alınarak sayısal değerler kullanılmamış ve katılımcıların kararlarının bu değerlerin küçük veya büyük olmasından etkilenmemesi sağlanmaya çalışılmıştır.

6. SONUÇ

Kullanılabilirlik değerlendirmeleri birçok farklı yöntemle yapılabilmektedir. Kullanıcılarla laboratuvar ortamında yapılan kullanılabilirlik testleri, gerçek kullanım verileri ve kullanım deneyiminin elde edilmesi açısından faydalı olsa da, kullanıcıların belirlenip davet edilmesi, testlerin tasarlanması, sonuçlarının analiz edilmesi için harcanan zaman ve kaynak, sorunların ortaya çıkarılması ve çözüm sunulması gibi adımları içeren maliyetli bir süreçtir. Bundan dolayı, bu çalışmada hızlı ve düşük maliyetli bir çözüm olarak kullanılabilirlik ilkelerine göre uzman gözden geçirmelerinin yapılabildiği bir sistem olarak KDS aracı tasarlanmıştır. Bu araçta, kullanıcıların değerlendirme yöntemi olarak klasik Likert tipi ölçeğe alternatif olarak sürgü ölçeği kullanılmış ve bulanık çıkarım yöntemi ile desteklenmiştir.

KDS aracında, referans alınan bir rehber ve ilke seti eşliğinde uzman kullanıcıların değerlendirme yapmalarına imkân tanınmıştır. Değerlendirme yöntemi olarak hem klasik Likert tipi ölçekte, hem de sürgü ölçeğinde çalışma gerçekleştirilmiştir. Sürgü ölçeği kullanımının radyo düğmesi gibi alternatiflere göre daha zor bir kullanıma sahip olması ve yanıt süresini artırması gibi dezavantajlarından bahseden çalışmalar literatür araştırması bölümünde bahsedilmiştir. Bununla birlikte, kullanıcıların Likert ölçeğinde ara değerler seçememesi ve belirlenen değerleri seçmeye zorlanması, aynı cevap seçilen iki farklı maddede aslında kullanıcı kararının tamamen aynı olmaması gibi dezavantajları olduğunu gösteren araştırmalara da yer verilmiştir.

Sürgü ölçeği kullanımının bireyler tarafından daha cazip ve eğlenceli bulunduğuna dair yapılan araştırmalardan bahsedilmiştir. Klasik Likert tipi ölçeklerde bireylerin kararlarını net olarak ifade edemedikleri ve bunun yerine sürgü ölçeğinin daha özgür ve esnek bir yapı sunduğu ifade edilmiştir. Bu tez çalışmasında, Likert tipi ölçek ile sürgü ölçeği kullanılarak uzman kullanıcılardan kullanılabilirlik ilkelerini değerlendirmeleri istenmiş ve çalışma sonunda elde edilen sonuçlar üzerinden kıyaslamalar yapılmıştır.

Bulanık çıkarım yöntemiyle desteklenen sürgü ölçeği ile yapılan değerlendirmeler sonucunda ortalama puanlar hesaplanmış ve 90'lı yıllardan itibaren güvenilir bir yöntem

olarak uygulanan sistem kullanılabilirlik ölçeği anketinden elde edilen puanlar ile karşılaştırılarak yöntemin güvenilirliği ölçülmüştür. Klasik değerlendirme yöntemi ile hesaplanan ortalama puanın bulanık değerlendirme ve SKÖ anketi puanından çok daha düşük çıktığı görülmüştür. Bu çalışmada önerilen bulanık değerlendirme yönteminin hem uzman kullanıcılar tarafından yapılan SKÖ anketi değerlendirme puanına çok yakın bir değer çıkardığı hem de tüm kullanıcıların yapmış olduğu SKÖ anketi değerlendirme puanına yakın bir değer çıkardığı görülmüştür.

KDS aracı üzerinde tanımlanmış olan KAMİS Rehber ilkeleri ve detaylı yönergeleri, web geliştirme alanında çalışan ancak kullanılabilirlik açısından çok fazla deneyimi olmayan kişilerin internet sitelerinin kullanılabilirliğini kolayca değerlendirmelerine yardımcı olmuştur. KAMİS Rehber ilkeleri ile değerlendirmek, kullanılabilirlik uzmanı olmayı gerektirmemektedir. Ancak bazı ilkeler web alanında teknik uzmanlığa sahip kişiler tarafından değerlendirilebilir niteliktedir. Bu çalışmada hazırlanan ve KAMİS Rehber ilkeleri ile internet sitelerinin değerlendirilmesini sağlayan KDS aracı; web geliştiricileri, web tasarımcıları gibi teknik deneyime sahip kişileri hedeflemektedir.

Bu araç ile basılı kitap ve internet sitesi üzerinden sunulan KAMİS Rehber ilkelerinin değerlendirilmesi kolaylaştırılmıştır. Aynı zamanda kullanılabilirlik çalışmalarının projelendirilerek planlanabilmesi ve hızlı bir şekilde değerlendirilmesi mümkün kılınmıştır. KDS aracının altyapısı aynı zamanda KAMİS Rehber dışında farklı ilke setlerinin eklenmesine ve yönetilebilmesine de imkân tanımaktadır. Bununla birlikte sistem üzerinden hem uzman değerlendirmelerinin yapılabilmesi, hem de son kullanıcılara yönelik bir anket değerlendirmesi olan SKÖ çalışmasının yapılabilmesi; bu iki yöntemle elde edilen puanların karşılaştırılabilmesini ve tutarlılığın teyit edilebilmesini sağlamıştır.

KDS aracı üzerinden yapılan çalışmada, bütün uzman kullanıcılar ilkeleri “Önem” ve “Uygunluk” parametreleri üzerinden değerlendirmişlerdir. Likert tipi ölçeğe göre yapılan klasik değerlendirme puanları ve sürgü-bulanık yönteminin kullanıldığı değerlendirme arasında tutarlılığın ölçülmesi amacıyla korelasyon hesaplaması yapılmıştır. Buna göre her iki yöntemde puanlanan “Önem” parametresi için korelasyon

katsayı değeri 0,79 çıkmıştır. “Uygunluk” parametresi için korelasyon katsayı değeri 0,55 çıkmıştır. Bu değerler her iki değerlendirme yönteminde puanlanan parametrelerin birbiri ile ilintili olduğunu göstermiştir.

Ayrıca “Önem” parametresi için 0,79 olarak, “Uygunluk” parametresi içinse 0,55 olarak hesaplanan korelasyon katsayısının istatistiki olarak anlamlı olduğunu %95 güven düzeyinde gösteren hipotez testi gerçekleştirilmiştir.

Klasik değerlendirme yönteminde tüm kullanıcıların değerlendirme ortalaması 70,84 puan olarak çıkmış, bulanık değerlendirme yönteminde 79,19 puan, SKÖ anketi değerlendirmelerinde ise ortalama puan olarak 85 çıkmıştır. Uzman değerlendirmesi yapmayan kullanıcıların SKÖ anketi puan ortalaması 87,24 çıkmış, uzman değerlendirmesi yapan kullanıcıların puan ortalaması ise 81,14 çıkmıştır. Bu sonucun bulanık değerlendirme yönteminden çıkan sonuca çok yakın olduğu görülmektedir. Uzman değerlendirme yapan kullanıcıların SKÖ anket sonucu 81,14 puanla, aynı kullanıcıların ortalama 79,19 olan bulanık değerlendirme sonuçlarını desteklemiştir.

İlke bazında uzman kullanıcılar tarafından yapılan değerlendirme sonrasında son puan, “Önem” ve “Uygunluk” parametreleri için analiz gerçekleştirilmiştir. Bu analize göre son puan ve “Önem” parametresi için her iki yönteme göre yapılan puanlamalar arasında korelasyon katsayısının yüksek olduğu görülmüştür.

“Uygunluk” parametresi için yapılan analizde diğer iki değerden farklı olarak korelasyon katsayılarının daha düşük olduğu görülmüştür. Yani, Likert tipi ölçek yöntemine ve sürgü-bulanık yöntemine göre yapılan değerlendirmelerde kullanıcıların birbiri ile çok tutarlı olmayan puanlar verdiği görülmüştür. Çalışma sonunda kısa bir anket yapılmış ve her iki yönteme göre ilişki düzeyinin düşük olması durumu bir soru olarak kullanıcılara yöneltilmiş ve kullanıcılardan olası sebeplerle ilgili olarak geribildirim alınmıştır. Bu durumun nedeni olarak en öne çıkan görüş; Likert tipi ölçek yönteminde “Önem” parametresinin rakamsal değerler olarak seçilmesi ve bunun algısal olarak daha belirgin bir yere sahip olması, dolayısıyla sürgü yönteminde de kullanıcıların benzer bir puanı vermiş olabileceği şeklinde belirtilmiştir. “Uygunluk”

parametresinin “Uygun değil” ve “Uygun” arasında değişen değerlere sahip olmasının, kullanıcıların algısında rakamsal bir değere doğrudan dönüştürülememesi ve dolayısıyla Likert tipi ölçek yönteminde kullanıcıların yapmış olduğu puanlamanın korelasyon katsayısını düşürdüğü ifade edilmiştir.

Olası bir gelecek çalışma için değerlendirilebilmek üzere, kullanılabilirlik ilkelerinin otomatik olarak kontrol edilmesine yönelik araştırmalar da yapılmıştır. Örneğin; Ek2 bölümünde sadece bir ilkenin otomasyonu için yapılan araştırma ve analiz çalışmasına yer verilmiştir. ISO 9241-151 standardı içerisinde “9.6.1 - Metnin Okunabilirliği” ilkesi açısından değerlendirme yapılmıştır. Standarda ait bu ilke KAMİS Rehber içerisinde “Yazı Görünümü” ilkesi ile ilişkilendirilmiştir. Ek2 bölümünde yer alan maddeler yazının sadece okunabilirlik yönü açısından otomatize edilebilirliğine yönelik değerlendirmelerdir. Ancak, bu çalışmada otomasyon özelliğine yer verilmesi halinde, kapsamın çok büyük bir ölçeğe taşacağı ve çok daha detaylı bir çalışma yapma ihtiyacı olacağı değerlendirilmiş ve bu konu, çalışmanın sonraki aşaması için bir araştırma konusu olarak önerilmiştir.

Uzmanlarla veya uzman olmayanlarla, kısacası insanlarla yapılan değerlendirmelerin hata veya öznellik içermesi ihtimali her zaman vardır. Buna karşı, özellikle uzman değerlendirme çalışmaları için yapay zekâ, makine öğrenmesi gibi ileri seviye mühendislik uygulamalarının kullanılması daha doğru sonuçlar elde edilmesi açısından önemlidir.

KAYNAKLAR

Anonim. (2008). Web Sitesi: https://tse.org.tr/IcerikDetay?ID=948, Erişim Tarihi: 07.03.2019.

Anonymous. (2012). Web Sitesi: https://www.w3.org/TR/WCAG21/, Erişim Tarihi: 27.03.2019.

Anonymous. (2013a). Web Sitesi: https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html, Erişim Tarihi: 07.03.2019.

Anonymous. (2013b). Web Sitesi: https://www.usability.gov/how-to-and-tools/methods/heuristic-evaluation.html, Erişim Tarihi: 07.03.2019.

Anonymous. (2013c). Web Sitesi: https://www.usability.gov/how-to-and-tools/methods/first-click-testing.html, Erişim Tarihi: 26.06.2019.

Anonymous. (2018). Web Sitesi: https://www.iso.org/standard/63500.html, Erişim Tarihi: 07.03.2019.

Arsoy, S., Kalipsiz, O. ve Öztürk, S. 2013. E-Devlet Web Siteleri İçin Rehber Tabanlı Kullanılabilirlik Değerlendirmesi, Yıldız Teknik Üniversitesi, İstanbul.

Bagchi, R. ve Davis, D. F. (2016). The role of numerosity in judgments and decision-making. *Current Opinion in Psychology*, *10*, 89–93.

Bakaev, M., Mamysheva, T. ve Gaedke, M. (2017). Current trends in automating usability evaluation of websites: Can you manage what you can't measure? *Proceedings - 2016 11th International Forum on Strategic Technology, IFOST 2016*, 510–514.

Bevan, N. (2005). *Guidelines and Standards for Web Usability, London, UK*.

BEVAN, N. (2001). International standards for HCI and usability. *International Journal of Human-Computer Studies*, *55*(4), 533–552.

Bevan, N., & Spinhof, L. (2007). Are Guidelines and Standards for Web Usability Comprehensive?, *Human-Computer Interaction. Interaction Design and Usability.* Lecture Notes in Computer Science. No: 4550, (407–419); Berlin.

Brooke, J. (1996). SUS—a quick and dirty usability scale. *Usability evaluation in industry*.189 (194), 4-7. Reading, UK.

Brooke, J. (2013). *SUS: A Retrospective*. Journal of Usability Studies. 8 (2), 29-40.

Cape, P. (2009). Slider Scales in Online Surveys. *CASRO Panel Conference*. New Orleans, US.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334.

de Vries, M., Holland, R. ve Witteman, C. (2008). Fitting decisions: Mood and intuitive versus deliberative decision strategies. *Cognition and Emotion*, *22*(5), 931–943.

Dobronte, A. (2012). LIkert Scales vs. Slider Scales in commercial market research. Web Sitesi: https://www.checkmarket.com/blog/likert-scales-slider-scales/, Erişim Tarihi: 01.04.2019.

Durmuş, S., & Çağıltay, K. (2012). Kamu Kurumu Web Siteleri ve Kullanılabilirlik. *E-devlet Kamu yonetimi ve teknoloji iliskisinde guncel gelismeler*. 293–322.

Fernandez, A., Insfran, E. ve Abrahão, S. (2011). Usability evaluation methods for the web: A systematic mapping study. *Information and Software Technology*, *53*(8), 789–817.

Funke, F. (2016). A Web Experiment Showing Negative Effects of Slider Scales Compared to Visual Analogue Scales and Radio Button Scales. *Social Science Computer Review*, *34*(2), 244–254.

Hub, M. ve Zatloukal, M. (2008). *Methodology of fuzzy usability evaluation of Information systems in Public administration. WSEAS Transactions on Information Science and Applications*. Pardubice, Czech Republic.

Kaya, A., Ozturk, R. ve Altin Gumussoy, C. (2019). Usability Measurement of Mobile Applications with System Usability Scale (SUS). *Industrial Engineering in the Big Data Era*, 389–400. Nevşehir, Turkey.

Likert, R. (1932). A Technique for the Measurement of Attitudes. *Archives of Psychology*, *22*, 5–55.

McLeod, S. (2008). Web Sitesi: https://www.simplypsychology.org/likert-scale.html, Erişim Tarihi: 28.03.2019.

Molich, R. ve Nielsen, J. (1990). Improving a human-computer dialogue. *Communications of the ACM*, *33*(3), 338–348.

Nielsen, J. (2000). Web Sitesi: https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/, Erişim Tarihi: 02.04.2019.

Nielsen, J. (2012). Web Sitesi: https://www.nngroup.com/articles/usability-101-introduction-to-usability/, Erişim Tarihi: 16.04.2019.

Nielsen, J. (1992). Finding usability problems through heuristic evaluation. *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '92*, 373–380. New York, USA.

Nielsen, J. (1994a). Enhancing the explanatory power of usability heuristics. *Conference companion on Human factors in computing systems - CHI '94*, 210. New York, USA.

Nielsen, J. (1994b). Usability inspection methods. *Conference companion on Human factors in computing systems - CHI '94*, 413–414. New York, USA.

Nielsen, J. ve Molich, R. (1990). Heuristic evaluation of user interfaces. *Proceedings of the SIGCHI conference on Human factors in computing systems Empowering people - CHI '90*, 249–256. New York, USA.

Pernice, K. ve Nielsen, J. (2009). *How to Conduct Eyetracking Studies*. Web Sitesi: http://www.nngroup.com/reports/how-to-conduct-eyetracking-studies/, Erişim Tarihi: 25.04.2019.

Roster, C. A., Lucianetti, L. ve Albaum, G. (2015). Exploring Slider vs. Categorical Response Formats in Web-Based Surveys. *Journal of Research Practice*, *11*(1), 1.

Toepoel, V. ve Funke, F. (2018). Sliders, visual analogue scales, or buttons: Influence of formats and scales in mobile and desktop surveys. *Mathematical Population Studies*, *25*(2), 112–122.

Wang, L., Bretschneider, S. ve Gant, J. (2005). Evaluating Web-Based E-Government Services with a Citizen-Centric Approach. *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, 129b-129b.

Wu, C.-I. (2012). HCI and Eye Tracking Technology for Learning Effect. *Procedia - Social and Behavioral Sciences*, *64(2012)*, 626–632.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, *8*(3), 338–353.

EKLER

EK 1 – Sistem Kullanılabilirlik Ölçeği (SKÖ) Anketi Soruları

EK 2 – Kullanılabilirlik İlkelerinin Otomasyonu

ÖZGEÇMİŞ

Adı Soyadı: Fatih TEKMEN

Doğum Yeri: Osmaniye

Doğum Tarihi: 29.05.1981

Medeni Hali: Evli

Yabancı Dili: İngilizce, Almanca

Eğitim Durumu (Kurum ve Yıl)

Lise: Osmaniye İHL (1999-2000)

Lisans: Viyana Teknik Üniversitesi, Bilgisayar Bilimleri Fakültesi, Yazılım ve Bilişim Mühendisliği (2008)

Çalıştığı Kurum/Kurumlar ve Yıl

Türk Telekom A.Ş. (2008 – 2014)

TÜBİTAK (2014 – 2016)

TÜBİTAK BİLGEM YTE (2016 – …)

Yayınlar (SCI) (Yayın yılları dikkate alınmalıdır)

Tekmen, F. ve Tanriover, O. 2017. Website usability evaluation with quickly applicable guidelines: An assessment of a government website, Global Journal of Computer Sciences: Theory and Research, *7*(1), 8, Ankara.

Tekmen, F. ve Tanriover, O. 2017. Kullanılabilirlik uzmanları için iyileştirilmiş web uygulaması değerlendirme sistemi, International Symposium on Industry 4.0 and Applications (ISIA 2017), 182–185, Ankara.