# AMAÇ

Bu araştırmanın amacı, büyük veri sistemlerinde ihtiyaca uygun kurumsal arama sistemi oluşturabilmektedir.

Arama motorları, kullanıcıların ihtiyaçlarına göre ilgili bilgileri kayda değer miktarda veri içerisinden sunan araçlardır. İşlenmesi gereken verinin büyüklüğü ve ilgili bilgileri kullanıcıya sunmak arama motorlarının iki ana problemini oluşturur. Arama motoru performansını artırmak için pek çok yaklaşım ve metod bulunmaktadır.

Bu ana amaç doğrultusunda aşağıdaki problemlere çözüm aranmıştır.

* Dünya üzerinde bulunan kurumsal veri arama sistemleri üzerine araştırmalar yapılıp açık kaynak kodlu sistemler belirlenmiştir.
* Belirlenen sistemlerin uyumlukları üzerinde araştırma yapılmıştır.
* Açık kaynaklar kullanılarak tüm kurumsal sistemlerinde küçük değişikliklerle ihtiyaca uygun bir şekilde çalışabilecek bir dosya yapısı oluşturulup farklı ihtiyaçlara uygun sistemlere uyum sağlaması sağlanacaktır.

# KONU

Araştırmanın konusu, büyük bilgi teknoloji alt yapısına sahip yapılarda büyük veri kümelerinin bulunması, bu sistemlerin içerisinde mevcut veritabanı sistemlerince indexlenemeyen freeformat bilgilerin istenilen süreden aranıp bulunamaması problemine çözüm olacak agent yapısının kurulması olacaktır.

Büyük yapılarda mevcut veritabanı sistemleri tarafından indexlene bilen birincil anahtar(primary key) alanlar ihtiyaca uygun olarak tekil(unique) oldukları için arama yapılabilir, uygun sürede arama sonucu alınabilir. Fakat ad, soyad, adres gibi bilgiler veritabanı sistemlerinde indexlenemeyen free text alanlardır, bu alanların indexlenemiyor olması mevcut sistemler tarafından arama yapılamamasına sebebiyet vermektedir. Bu alanlar içerisinden veri kümeleri araması yapılabilmesi için bu alanların ihtiyaca uygun bir şekilde indexlenebiliyor olması gerekmektedir.

Milyonlarca müşteri datası olan bir sistem düşündüğümüzde her müşteriye ait farklı sıralamalarda birden fazla isim, soyad ve birden fazla adres olabilir. Bu şekilde datanın büyüklüğü mevcut durumunun çok fazla katına çıkmaktadır, ayrıca bu bilgiler farklı farklı veri tabanı sistemlerinde olabilir, bu durum arama yapısını ve aranacak verinin tüm sistemlerde tekil indexlenmesini zorlaştırmıştır. Çözüm için tekil bir sistemde tüm verinin indexlenmesi gerekmektedir.

Araştırma konusu olarak indexlenemeyen veri kümelerinde açık kaynaklı arama sistemi yardımıyla tüm sistemlere entegre edilecek bir yapı kurulmasıdır.

# LİTERATÜR ARAŞTIRMASI

Bilgi, bilişim sistemlerinin tüm süreçlerinde ihtiyaç duyulan bir kaynaktır, bilgiye erişim ne kadar hızlı olursa kullanıcı deneyim ve bilişim uygulamalarının kullanışı artmaktadır.

Bilgiyi aramak için kullanılan ilk yöntem veritabanında aranacak metinlerin indekslenmesidir. Birebir eşleşen yapılarda (örneğin ID, isim) yapılan aramalarda kullanılan yaygın yöntemdir. Uzun metin içeriklerinde arama yapılmak istenildiğinde bu yöntem kullanılamamaktadır.

Bilişim sistemlerinde kurumların büyüklüğüne göre arama kümesinin büyüklüğü ve arama kümesini büyüme hızı değişmektedir. Milyonlarca datanın olduğu sistemler, arama yapısının karmaşıklığını artırmaktadır, buna ek olarak online olarak sürekli artan data yapıları, mevcut arama sistemlerini daha da karmaşık hale getirmektedir. Kurumsal arama motorlarının temel amacı büyük veri yapılarında, birbirinden bağımsız sistemlerde ortak arama sisteminin oluşması ve anlık olarak beslenen verilerin sisteme anlık olarak dahil edilmesi için çözüm sunmaktadır.