BİÇİMSEL DİLLER VE OTOMATA TEORİSİ

HAFTA 1

GİRİŞ, TEMEL KAVRAMLAR

DR. ÖĞR. ÜYESİ SİNEM AKYOL
sakyol@firat.edu.tr

Dersin Amacı

Biçimsel dil teorisinin içeriğini oluşturan otomatlar, diller ve gramerler konusunu öğretmek, sonlu otomata teorisi, alta bastırmalı otomata teorisi ve Turing makineleri hakkında bilgi sahibi olmayı sağlamak ve biçimsel düşünme yeteneğini kazandırmaktır.

Dersin İçeriği

- Otomata teorisine giriş, alfabe, sembol, kelime ve dil kavramları
- Özyinelemeli tanımlar
- Düzenli ifadeler
- Sonlu Otomata
- Geçiş çizgeleri
- Çıktılı Sonlu Otomata
- Düzenli ve düzenli olmayan diller
- Durumdan bağımsız dilbilgileri
- Alta bastırmalı Otomata Teorisi
- Yığıtlı Otomata
- Ayrıştırma İşlemleri
- Turing Makineleri

Ders Kitabı Malzemesi, Önerilen Kaynaklar

- Introduction to Automata Theory, Languages, and Computation (2nd Edition), John E. Hopcroft, Rajeev Motwani, Jeffrey D. Ullman, lAddison Wesley, 2000.
- Elements of the Theory of Computation (2nd Edition), Harry R. Lewis, Christos H. Papadimitriou, Prentice Hall, 1998.
- Özdevinirler Kuramı ve Bçimsel Diller, Ünal Yarımağan, Bıcaklar Kitapevi, 2003
- Ders Notları

Otomata Teorisi (Automata Theory)

- Otomata teorisi, soyut hesaplama aygıtlarının veya makinelerinin incelenmesidir.
- A. Turing, günümüz bilgisayarlarının tüm özelliklerine sahip, en azından hesaplayabildiği kadarıyla soyut bir makine üzerinde çalıştı.
- Turing'in amacı, bir hesaplama makinesinin yapabilecekleriyle sonuçlarının neler yapamayacağı arasındaki sınırı tam olarak tanımlamaktı, sadece soyut Turing makinelerine değil, bugünün gerçek makinelerine de.

Alan Turing

Hedeflenen Kazanımlar

- Problemleri soyutlayabilme, basitçe ifade edebilme
- Sınırlı kapasite ile çalışabilme deneyimi elde etme.
- İlk bilgisayarların nasıl çalıştığı hakkında bilgi sahibi olma.

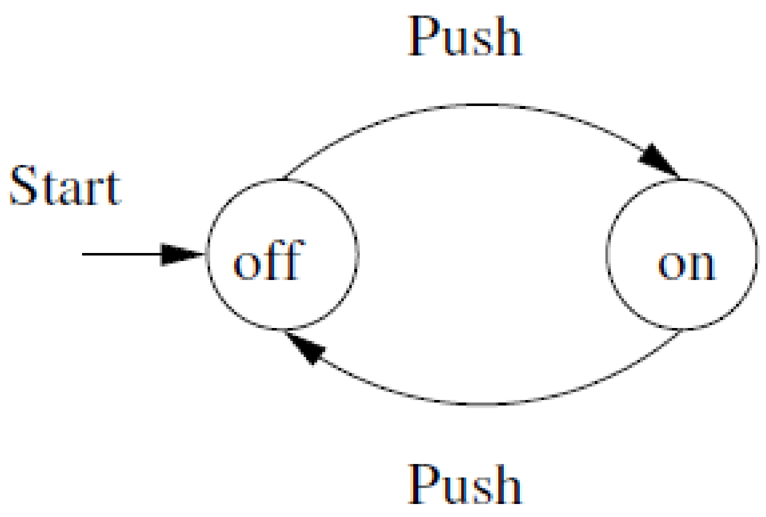
Sonlu Otomata (Finite Automata)

- 40’larda ve 50’lerde çalışılmış.
- Başlangıçta beyin fonksiyonunu modellemek için önerilmiştir.
- Daha sonra, diğer birçok amaç için son derece yararlı olduğu ortaya çıktı.
- 1969’da Cook, Turing’in neyin hesaplanıp neyin hesaplanamayacağına ilişkin çalışmasını genişletti.
- Moore Yasası (Moore’s Law)

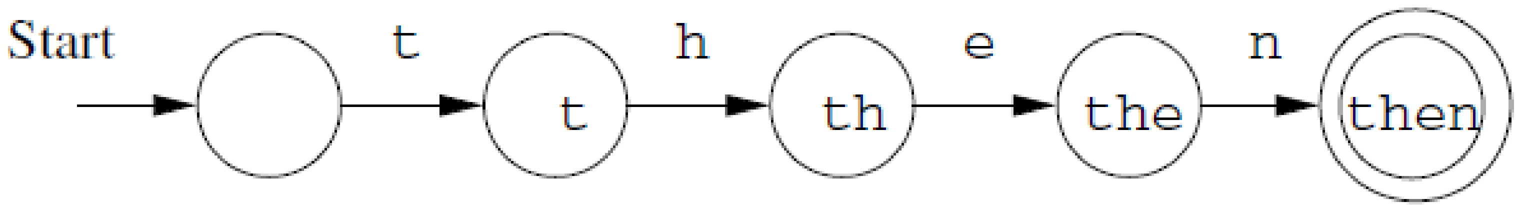
Neden Sonlu Otomata?

- Dijital devrelerin davranışını tasarlama ve kontrol etmek için yazılım.
- Girdi metnini mantıksal birimlere ayıran derleyici bileşeni olan tipik bir derleyicinin «sözcüksel analizi», tanımlayıcılar, anahtar kelimeler ve noktalama işaretleri gibi.
- Kelime cümlelerinin veya diğer kalıpların oluşumlarını bulmak için büyük miktarda metin içeren Web sayfalarının koleksiyonları gibi yazılımlar.
- Güvenli bilgi alışverişi için iletişim protokolleri veya protokoller gibi sınırlı sayıda farklı duruma sahip olan her türlü sistemi doğrulama yazılımı

- "açık" durumda mı yoksa "kapalı" durumda mı olduğunu hatırlayan, en basit sonlu otomata



- «then» i tanıyan otomata



Otomata Teorisi neden çalışılmalı?

- Bir bilgisayar ne yapabilir? Bu çalışmaya "karar verilebilirlik (decidability)" denir ve bilgisayar tarafından çözülebilecek problemlere "karar verilebilir (decidable)" denir.

- Bir bilgisayar verimli bir şekilde ne yapabilir? Bu çalışmaya "kararsızlık" denir ve girdi boyutunun yavaş yavaş büyüyen bir fonksiyonundan daha fazla zaman kullanmadan bir bilgisayar tarafından çözülebilecek sorunlara "izlenebilir" denir.

Temel Kavramlar

- **Kümeler – Tanım**
- Bir küme, sıralanmamış nesne topluluğudur.
- $S = \{a, b, c\}$
- $a, b\ ve\ c,$ S kümesinin elemanlarıdır.
- Bir küme içindeki elemanların birbirleriyle ilişkisi yoktur.
- $S_1 = \{1, 2, 3\}$
- $S_2 = \{red, farmhous,\ \pi, -32\}$

- **Kümeler – Tanım**
- Kümeler eleman olarak diğer kümeleri içerebilirler.
- $S_1 = \{3, \{3, 4\}, \{4, \{5,6\}\}\}$
- $S_2 = \left\{\{1, 2\}, \left\{\{3\}\right\}\right\}$
- Kümeler çift içermiyor
- $NotASet = \{4,2,4,5\}$

Kümeler ve Bağıntılar

- Bir küme nesnelerden oluşur
  - *L = {a, b, c, d}*
  - *a, b, c, d* kümenin elemanları veya üyeleridir
  - *c ∈ L, k ∉ L şeklinde ifade edilir.*
- Elemanların sırası ve tekrarı önemli değildir
  - *{üzüm, kiraz, üzüm}* ile *{üzüm, kiraz}* aynıdır
  - *{c, k, h}*, *{h, c, k}* ve *{k, c, h}* aynıdır
- Elemanların birbiri ile ilişkili olması gerekmez
  - {3,üzüm,{c,mavi}}
  - biri kendi içinde küme olan üç elemana sahip bir kümedir.

Kümeler ve Bağıntılar

- Empty ve singleton
  - Bir elemana sahip küme *singleton*, hiç elemanı olmayan küme *empty* olarak adlandırılır.
  - *{1}, {kiraz}* singleton
  - *{ }*, Ø empty set
- Sonsuz küme
  - *N = {0, 1, 2, 3, ...}* doğal sayılar kümesi
- Kümeler özellikleriyle de tanımlanabilir
  - *G = {x | x ∈ I and x is greater than 2}*
  - *O = {x | x ∈ N and x is not divisible by 2} tek sayılar kümesi*

Kümeler ve Bağıntılar

- Altküme
  - $A \subseteq B$, *A* kümesi *B* kümesinin altkümesi (*A = B olabilir)*
  - $A \subset B$, *A* kümesi *B* kümesinin *proper* altkümesi ($A \neq B$)
- Union (Birleşim)
  - $A \cup B = \{x \mid x \in A \text{ or } x \in B\}$
  - $A = \{0, 1, 2\}$, $B = \{7, 8, 9\}$ □ $A \cup B = \{0, 1, 2, 7, 8, 9\}$
- Intersection (Kesişim)
  - $A \cap B = \{x \mid x \in A \text{ and } x \in B\}$
  - $A = \{0, 1, 2\}$, $B = \{8, 9, 0\}$ □ $A \cap B = \{0\}$

Kümeler ve Bağıntılar

- Difference (Fark)
  - $A - B = \{x \mid x \in A \text{ and } x \notin B\}$
  - $A = \{1, 3, 9\}$, $B = \{3, 5, 7\}$ □ $A - B = \{1, 9\}$
- Disjoint (Bağımsız Kümeler)
  - $A \cap B = \{\ \}, \varnothing$

Kümeler ve Bağıntılar

Küme işlemleri

- *Idempotency* $A \cup A = A$

  $A \cap A = A$

- *Commutativity* $A \cup B = B \cup A$

  (Değişme) $A \cap B = B \cap A$

- *Associativity* $(A \cup B) \cup C = A \cup (B \cup C)$

  (İlişkisellik) $(A \cap B) \cap C = A \cap (B \cap C)$

Kümeler ve Bağıntılar

- *Distributivity* $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$
  (Dağılma) $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$

- *Absorption* $(A \cup B) \cap A = A$
  $(A \cap B) \cup A = A$

- *DeMorgan's laws* $A - (B \cup C) = (A - B) \cap (A - C)$
  $A - (B \cap C) = (A - B) \cup (A - C)$

$S = \{\{a, b\}, \{b, c\}, \{b, c, d\}\}$
$A = \{a, b, c, d\}$

Kümeler ve Bağıntılar

- **Birden fazla kümede birleşim**

$\cup S = \{x: x \in P \text{ for some set } P \in S\}$

$\cup S = \{a, b, c, d\}$

- **Birden fazla kümede kesişim**

$\cap S = \{x: x \in P \text{ for each set } P \in S\}$

$\cap S = \{b\}$

- **Power set**

Bir kümenin boş kümede dahil tüm altkümeleri

2A, A kümesinin power kümesi

$A = \{c, d\}$ ise $2\{c, d\} = \{\{c, d\}, \{c\}, \{d\}, \emptyset\}$

- **Partition**

Π power kümenin altkümesidir, boş kümeyi içermez ve A kümesinin her

  - elemanını sadece bir kez bulundurur
  - Π içindeki her eleman boş kümeden farklıdır
  - Π içindeki farklı elemanlar disjoint kümedir
  - $\cup\Pi = A$
  - $\{\{a, b\}, \{c\}, \{d\}\}$ partition, $\{\{b, c\}, \{c, d\}\}$ partition değil

Kümeler ve Bağıntılar

- **Binary relation**

A ve B kümeleri arasında binary relation AxB ‘nin altkümesidir

Örnek:

{2, 3} ve *{a, c}* kümeleri arasında *{(2, a), (3, a)}* bir binary relation olarak tanımlanır.

$\{(i, j): i, j \in N \text{ ve } i < j\}$ küçüktür ilişkisi olup NxN’nin altkümesidir

{(1, 2), (1, 3), (2, 6), ...} şeklinde sonsuz elemana sahiptir

Kümeler ve Bağıntılar

- **Tuples and relations**

(a1, a2, a3, ..., an) ordered tuple olarak adlandırılır (n-tuple)

n= 2 için ordered pairs, n= 3 için ordered triples

n= 4 için quadruples, n= 5 için quintuples

n= 1 için unary relation n= 2 için binary relation

n= 3 için ternary relation n-ary relation

Kümeler ve Bağıntılar

- **Function**

A ve *B* kümeleri arasında bir fonksiyon, binary relation *R = (a, b)*'dir ve her *a* ∊ *A* için kesinlikle sadece bir ordered pair vardır.

R1={(x,y) : x ϵ C, y ϵ S, and x is a city in state y} → fonksiyon

R2={(x,y) : x ϵ S, y ϵ C, and y is a city in state } → fonksiyon değil

f : *A* → *B*, *A*' dan *B*' ye tanımlanmış *f* fonksiyonu

Kümeler ve Bağıntılar

- **Domain ve Image**

$f: A \rightarrow B,\ a \in A$

A domain olarak adlandırılır

$f(a)$ image olarak adlandırılır ve her a için unique değerdir

- **Arguments ve Value**

$f : A1 \times A2 \times \ldots \times An \rightarrow B$ fonksiyon ise $f(a1, a2, \ldots, an) = b$ şeklinde gösterilir ve $ai \in Ai$, $i = 1, \ldots, n$ ve $b \in B$'dir.

Burada ai arguments ve b ise value olarak adlandırılır.

Otomata Teorisinin Temel Kavramları

- **Alfabe:** Bir semboller kümesi
- **Dizeler (Strings):** Alfabe sembollerinden oluşan liste
- **Dil:** Aynı alfabe harflerinden oluşan bir dizi

Alfabe

- Geleneksel olarak ∑ ile gösterilir.
- Yaygın alfabe örnekleri:
  - ∑={0, 1}, ikili alfabe
  - ∑={a, b, c, .., z} tüm küçük harflerin kümesi
  - ASCII karakterlerin kümesi

Dizeler (Strings)

- Bir dize (veya bazen kelime) bazı alfabelerden seçilen sonlu bir sembol dizisidir.

  01101, ∑={0, 1} ikili alfabesinden bir dizedir
- Boş dize, sıfır sembol oluşumuna sahip dizedir. ε olarak adlandırılır.
- dizgideki sembollerin (yani sembollerin sayısı) konumlarına dize uzunluğu denir.

  -'00110101'in uzunluğu 8.
- ω dizgisinin uzunluğu için standart gösterim | ω |

  |00110101|=8

  | ε |=0

Alfabelerin güçleri

- Σ bir alfabe iken Σ^k, Σ alfabesinin harflerinden oluşan k uzunluğundaki kelimelerin kümesidir.

  ör. Σ=0,1iken Σ^1=0,1,Σ^2=00,01,10,11,…

Not: Tüm Σ alfabeleri için $\Sigma^0=\varepsilon$. Yani yalnızca boş kelimeyi içerir.

- $\Sigma*$: Σ alfabesinin harfleri kullanılarak oluşturulabilecek bütün kelimelerin kümesidir:

$$\Sigma* = \Sigma^0 \cup \Sigma^1 \cup \Sigma^2 \cup \ldots$$

  ör. Σ=0,1için Σ^* =ε,0,1,00,11,01,10,000,111,010,….

- Dil: Bir Σ alfabesi üzerine bir dil Σ^*'ın bir alt kümesidir. L ile gösterilir.

ör. L=0,00,001,0000,00001,….→Σ={0,1} alfabesi üzerine 0 ile başlayan kelimelerin dili.

ör. $L=\{\varepsilon,$ 01, 10, 0011, 0101, 1001,….}→Σ=0,1alfabesi üzerine içerdiği toplam 0 sayısı içerdiği toplam 1 sayısına eşit olan kelimelerin dili.

ör. Türkçe dili Σ={a, b, c, ç….y, z} üzerine kelimelerin kümesi.

- Dillerin bitiştirilmesi (concatenation of languages) : Kelimelerin bitiştirilmesi kavramı diller için de genişletilebilir. L_1 ve L_2 iki dil iken bunların bitiştirilmesi

$$L_1L_2=\{xy \mid x\in L_1, y\in L_2\}$$

ör. $L_1=\{0,01\}, L_2=\{\varepsilon,b,bb\}$ iken $L_1L_2=\{0,0b,0bb,01,01b,01bb\}$

$0\varepsilon=0$

TEŞEKKÜRLER…