Sifat-sifat Bahasa-bahasa Reguler

Kuliah Teori Bahasa dan Automata
Program Studi Ilmu Komputer
Fasilkom UI

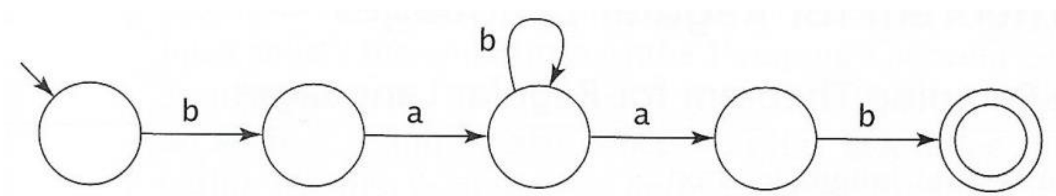
Prepared by:
Suryana Setiawan

Bahasa Reguler dan nonreguler

- Pembahasan hingga saat ini, suatu bahasa L adalah bahasa reguler dengan menunjukkan untuk L tsb:
  - dapat dispesifikasikan dengan ekspresi regular, atau
  - dapat dibuat suatu FSM, atau
  - dapat ditemukan dalam jumlah berhingga kelas-kelas ekivalen dari relasi $\approx_L$
  - dapat dinyatakan dengan suatu grammar reguler
  - sifat-sifat closure untuk bahasa reguler berlaku.
- Untuk bahasa nonreguler L, bagaimana caranya membuktikan L bukan bahasa reguler (bahwa hal-hal di atas tidak ada untuk L)?

Observasi Bahasa Reguler

- Setiap bahasa reguler L dapat diterima oleh suatu DFSM M (Ingat: $|K|$ **berhingga**). Jika L adalah **tak berhingga** maka
  - Pasti sekurangnya terdapat satu loop di dalam M (pigeonhole principle).
- Setiap $w \in L$, $|w| \geq |K|$, memiliki satu atau lebih pola berulang karena dengan adanya substring y ($y \neq \varepsilon$, dan $w = xy^q z$, $q \geq 0$) yang membawa M masuk dalam loop.
  - *Contoh:* $w = babab$ dengan $x = ba$, $y = b$, $z = ab$



Observasi Bahasa Nonreguler

- Observasi dalam slide sebelumnya tidak dapat digunakan untuk memeriksa suatu bahasa L yang sembarang yang tidak diketahui apakah ada FSM untuk menerimanya.
  - Kalau pasti ada, pasti bahasa reguler!
  - Kalau pasti tidak ada, pasti bukan bahasa reguler!
- Jika tidak pasti? Asumsikan terdapat suatu FSM *M* (tapi entah seperti apa!) dengan jumlah status *k*,
  - Jika observasi itu berlaku, maka itu bahasa reguler.
  - Jika observasi tidak berlaku, bukan bahasa reguler.

Teorema Pumping

- (Istilah lain: "Pumping Lemma") Jika L adalah bahasa reguler, maka:
  - $\exists k \geq 1$ (
    $\forall\ w \in L$, dimana $|w| \geq k$ (
    $\exists x, y, z$ (
    $w = xyz$, $|xy| \leq k$, $y \neq \varepsilon$, dan,
    $\forall q \geq 0\ (xy^q z \in L)$
)
)
)

Bukti Teorema Pumping

- Dari observasi sebelumnya dengan
  - $k = |K|$,
  - string w dengan panjang k atau lebih,
  - y adalah substring yang membawa M melalui loop,
  - x prefiks dari w sebelum y
  - z sufiks dari w setelah y
- Maka berlaku
  - $|xy| \le k$: tanpa melalui loop maka string terpanjang kurang dari $|K|$, jika tepat ada k simbol, loop dilalui satu kali.
  - $y \ne \varepsilon$: karena M deterministik, tidak ada loop yang dilalui dengan ε
  - $\forall q \ge 0\ (xy^q z \in L)$: untuk q = 0, M tidak melalui loop, untuk q > 0, maka M melalui loop sekurangnya sekali.

Teorema Pumping untuk Nonreguler

- Untuk suatu bahasa yang diketahui reguler maka teorema pasti terbukti, untuk suatu bahasa yang diketahui nonreguler teorema pasti tidak terbukti.
- Untuk bahasa L yang tidak diketahui, pembuktian dengan kontradiksi (*proof by contradiction*)!
  - "Jika L reguler, maka harusnya memiliki sifat-sifat tertentu. Nyatanya ia tidak memiliki sifat-sifat itu. Jadi, ia bukan reguler."

Contoh: Bahasa A^nB^n

- Diketahui L adalah $A^nB^n = \{a^nb^n : n \geq 0\}$.
  - "Apabila L reguler harusnya terdapat k sehingga setiap string w, dengan $|w| \geq k$, memenuhi kondisi-kondisi dalam teorema pumping."
  - Kontradiksi ditunjukkan dengan mengambil suatu w yang tidak memenuhinya!
- Periksa misalkan untuk $w = a^kb^k$, sehingga $|w| = 2k \geq k$
- Adakah x, y, dan z sehingga $w = xyz$ dimana $|xy| \leq k$, $y \neq \varepsilon$, dan, $\forall q \geq 0$ selalu berlaku $xy^qz \in L$?
  - Coba $y = a^p$, $x = a^{k-p}$ (memenuhi $|xy| \leq k$) dan $z = b^k$, maka $a^{k-p}(a^p)^qb^k \notin L$, untuk $q \geq 0$ dan $q \neq 1$ => kontradiktif. (Kemungkinan yang lain juga membawa ke kontradiktif!)
- Coba w yang lain???

Contoh: Bahasa A^nB^n (coba *w* lain)

- Mencoba w = $a^{\lceil k/2 \rceil} b^{\lceil k/2 \rceil}$ dengan $|w| = k$ jika k bil genap atau $|w| = k+1$ jika k bil ganjil.
- Jika $|y| = p$, berarti y dapat berupa a^p atau $a^{p-j}b^j$ atau b^p
  - Kasus $y = a^p$, dengan $q = 2$, menghasilkan $a^{\lceil k/2 \rceil - p} (a^p)^2 b^{\lceil k/2 \rceil} = a^{\lceil k/2 \rceil + p}\, b^{\lceil k/2 \rceil} \notin L$
  - Kasus $y = a^{p-j}b^j$, dengan $q = 2$, menghasilkan $a^{\lceil k/2 \rceil - p + j} (a^{p-j}b^j)^2 b^{\lceil k/2 \rceil} = a^{\lceil k/2 \rceil}\, b^j a^{p-j} b^{\lceil k/2 \rceil} \notin L$
  - Kasus $y = b^p$, dengan $q = 2$, menghasilkan $a^{\lceil k/2 \rceil} (b^p)^2 b^{\lceil k/2 \rceil - p} = a^{\lceil k/2 \rceil}\, b^{\lceil k/2 \rceil + p} \notin L$
- Mencoba berbagai w akan tetap menghasilkan kontradiksi!

Contoh: untuk Bal

- Bal = {w $\in$ { (,) }* : tanda kurung berpasangan}
- Kita ingin mendapatkan string *w* sehingga teorema sukses menunjukkan L reguler.
  - Coba dengan $w = (^k)^k$, $|w| = 2k$
- Akan sama halnya dengan pembuktian A^nB^n, mengakibatkan banyaknya simbol "(" berbeda dengan simbol ")".

Contoh: untuk PalEven

- PalEven = $\{ww^R : \in \{a, b\}^* \}$; bahasa palindrom dengan panjang string bil genap.
  - Kita ingin mendapatkan string w sehingga teorema sukses menunjukkan L reguler.
- Coba dengan $w = a^k b^k b^k a^k$, $|w| = 4k$
  - Akan sama halnya dengan pembuktian $A^n B^n$, mengakibatkan banyaknya deretan simbol a terkiri berbeda dengan deretan simbol a terkanan.
- Coba dengan $w = a^{\lceil k/2 \rceil} b^{\lceil k/2 \rceil} b^{\lceil k/2 \rceil} a^{\lceil k/2 \rceil}$, $|w| \geq 2k$
  - Akan sama halnya dengan pembuktian $A^n B^n$, mengakibatkan ruas kiri berbeda dengan ruas kanan.
- dst

Contoh: Bahasa A^nB^m, n > m

- Mencoba w = $a^{k+1} b^k$ dengan $|w| = 2k+1$.
- Jika $|y| = 1$, berarti $x = a^k$ dan $z = b^k$
  - Disini tentunya jangan mengambil q > 0 karena dapat menghasilkan $a^k a^q b^k = a^{k+q} b^k \in L$
  - Maka ambil $q = 0$, menghasilkan $a^k a^0 b^k = a^k b^k \notin L$ (kontradiksi).

Contoh: Bahasa A^n dengan n Bil. Prima

- L adalah $\text{Prime}_a = \{a^n : n \text{ bilangan prima}\}$
- Dengan diberikan suatu *k* bukan prima, mengambil *j* bilangan prima terkecil yang lebih besar dari *k*, sehingga $|w| = j > k$.
- Ambil suatu *y* yang tidak kosong, dan *x* & *z* sisanya, sehingga $xy \leq k$ maka berikutnya adalah menguji $\forall q \geq 0$, $a^{|x|}(a^{|y|})^q a^{|z|} = a^{|x|+|z|+p|y|}$ apakah selalu $\in L$,
  - Berarti memeriksa $|x| + |z| + q\,|y|$ apakah selalu bilangan prima?
- Namun, $|x| + |z| = q$, dapat mengakibatkan sbb.
  - $|x| + |z| + q\,|y| = q + q\,|y| = q(|y| + 1) \Rightarrow$ bukan bil prima karena berfaktor $|xz|$ & $(|y|+1)$ dan keduanya > 1.

Kesimpulan Pembuktian Teorema Pumping u/ Nonreguler

- Ada dua harga yang perlu dicari untuk memenuhi properti dari teorema: *w* dan *q*.
- Dalam menentukan w, pilihlah yang:
  - bagian dari *L* yang membuat *L* tidak reguler.
  - “only barely” dalam *L*.
  - Region awal dengan panjang $\geq k$ yang se-homogen mungkin
- Dalam menentukan *q*, pilihlah yang:
  - Mencoba harga 0 atau 2.
  - Jika tidak berhasil, temukan harga yang mungkin berhasil dengan mempelajari *L*.

Teorema Pumping u/ Reguler?

- Tujuan teorema untuk menunjukkan dengan pasti suatu bahasa yang tidak memiliki sifat pumping, adalah nonreguler.
- Tetapi untuk mengetahui apa yang terjadi jika teorema digunakan untuk suatu bahasa reguler, akan dibahas dl slide berikutnya untuk bahasa A^nB^m yang sudah jelas bahasa reguler.

Contoh: untuk Bahasa A^nB^m

- Diketahui L adalah $A^nB^m = \{a^nb^m : m, n \geq 0\}$.
  - Disini kita ingin mencari w yang akan menyebabkan teorema gagal menunjukkannya bahasa reguler.
- Periksa untuk $w = a^kb^k$, sehingga $|w| = 2k \geq k$, lalu:

  $y = a$,

  $x = a^{k-1}$ (untuk memenuhi $|xy| \leq k$) dan

  $z = b^k$,

  maka $a^{k-1}a^qb^k = a^{k+(q-1)}b^k$ yang selalu $\in L$, untuk $q \geq 0$.
- Silahkan mendapatkan w lain untuk mengujinya!

Memahami Teorema Pumping melalui suatu permainan

1. Lawan: mengajukan L untuk diperiksa
2. Anda: harus membuktikan bhw L nonreguler
3. Lawan: berasumsi adanya FSM dengan jumlah status k (k boleh suatu bilangan atau tetap general sebagai k)
4. Anda: menemukan suatu $w \in L$ yang tepat, $|w| \geq k$
5. Lawan: menemukan substring-substring x, y, z, dimana $w = xyz$, $|xy| \leq k$, dan $|y| \geq 1$, untuk dapat membatalkan usaha anda
6. Anda: menemukan suatu q sehingga $xy^qz \notin L$. Jika gagal, ulangi langkah 4 hingga memang disadari ternyata memang reguler!

Memanfaatkan sifat Closure

- Hasil operasi yang bersifat closue L dengan suatu bahasa reguler L_R, jika L reguler maka hasil operasi L' juga harus reguler. Namun, ternyata L' bukan reguler, maka juga berarti L bukan reguler.
- Operasi irisan:

  $L = \{w \in \{a, b\}^*: \#_a(w) = \#_b(w)\}$

  Karena $L \cap A^nB^m = A^nB^n$, dan A^nB^n nonreguler, maka L juga nonreguler.
- Operasi komplemen:

  $L = \{a^ib^j: i, j \geq 0 \text{ dan } i \neq j\}$

  Karena $\neg L \cap A^nB^m = A^nB^n$, dan A^nB^n nonreguler, maka $\neg L \cap A^nB^m$ juga nonreguler, lalu L nonreguler karena $\neg L$ nonreguler.

Terkadang Manfaat Sifat Closure memang Diperlukan!

- $L = \{a^i b^j c^k : i, j, k \geq 0$ dan (jika i=1, maka $j = k$)$\}$
- Jika lansung diuji dengan teorema pumping, maka setiap string w, dengan $|w| > 0$ selalu bisa dipumping.
  - Jika $i = 0$ (yaitu $w = b^j c^k$), untuk $j \neq 0$, dengan $y = b$ (otherwise $k \neq 0$, dengan $y = c$), maka $\forall q \geq 0$, $xy^q z \in L$
  - Jika $i = 1$ (yaitu $w = ab^j c^k$), dengan $y = a$, maka $\forall q \geq 0$, $xy^q z \in L$
  - Jika $i = 2$ (yaitu $w = aab^j c^k$), dengan $y = aa$, maka $\forall q \geq 0$, $xy^q z \in L$
  - Jika $i > 2$, dengan $y = a$, maka $\forall q \geq 0$, $xy^q z \in L$
- Namun, L bukan reguler!

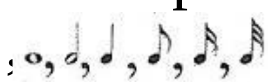
Terkadang Manfaat Sifat Closure memang Diperlukan! (lanjutan)

- $L = \{a^i b^j c^k : i, j, k \geq 0$ dan (jika i=1, maka $j = k$)$\}$
  - Sifat nonreguler L akibat string-string dengan i =1
- Perhatikan bahwa

  $$L' = L \cap \{ab^j c^k : j, k \geq 0\} = \{ab^j c^j : j \geq 0\}$$

  - $\{ab^j c^k : j, k \geq 0\}$ adalah bahasa reguler dengan ekspresi reguler ab^*c^*
  - Jika $\{ab^j c^j : j \geq 0\}$ reguler maka L reguler menurut sifat closure.
  - Tetapi $\{ab^j c^j : j \geq 0\}$ dengan teorema pumping dapat dibuktikan tidak reguler.
- Cara lain: jika L reguler, L^R juga reguler. Pemeriksaan dengan teorema pumping pada L^R dapat segera membuktikan L^R bukan reguler, maka juga L bukan reguler.

Problem-problem Dunia Nyata

- Bahasa-bahasa yang telah dibahas cukup sederhana sehingga sifat-sifat reguler/nonreguler cukup eksplisit.
- Bahasa-bahasa dari dunia nyata ternyata "tidak sesederhana itu."
- Contoh, apakah L reguler?
  - $L = \{w \in \{0, 1, \ldots, 7\}^*$: w representasi oktal dari bilangan bulat non-negatif yang habis dibagi 7$\}$ $= \{0, 7, 16, 25, \ldots\}$.
  - $L = \{w \in \{$  $\}^*$: w merepresentasikan suatu lagu dengan ketukan 4/4$\}$,  adalah simbol-simbol not dengan 4 ketuk, 2 ketuk, 1 ketuk, ½ ketuk, ¼ ketuk dan 1/8 ketuk, masing-masing.