

Automata and Formal Languages (2)

Email: christian.urban at kcl.ac.uk
Office: SI.27 (1st floor Strand Building)
Slides: KEATS

Languages

A **language** is a set of strings.

A **regular expression** specifies a set of strings, or language.

Strings

Different ways of writing strings:

”bello”

[b,e,l,l,o]

b::e::l::l::o::Nl

””

[]

Nl

Strings

Different ways of writing strings:

"bello" $[b, e, l, l, o]$ $b :: e :: l :: l :: o :: Nil$

"" $[]$ Nil

The concatenation operation on strings and sets of strings:

"foo" @ "bar" = "foobar"

$A @ B \stackrel{\text{def}}{=} \{s_1 @ s_2 \mid s_1 \in A \wedge s_2 \in B\}$

Regular Expressions

Their inductive definition:

$r ::=$	\emptyset	null
	ϵ	empty string / "" / []
	c	character
	$r_1 \cdot r_2$	sequence
	$r_1 + r_2$	alternative / choice
	r^*	star (zero or more)

Re

Their indu

```
abstract class Rexp  
  
case object NULL extends Rexp  
case object EMPTY extends Rexp  
case class CHAR(c: Char) extends Rexp  
case class ALT(r1: Rexp, r2: Rexp) extends Rexp  
case class SEQ(r1: Rexp, r2: Rexp) extends Rexp  
case class STAR(r: Rexp) extends Rexp
```

$r ::= \emptyset$	null
ϵ	empty string / "" / []
c	character
$r_1 \cdot r_2$	sequence
$r_1 + r_2$	alternative / choice
r^*	star (zero or more)

The Meaning of a Regular Expression

$$L(\emptyset) \stackrel{\text{def}}{=} \emptyset$$

$$L(\epsilon) \stackrel{\text{def}}{=} \{\epsilon\}$$

$$L(c) \stackrel{\text{def}}{=} \{c\}$$

$$L(r_1 + r_2) \stackrel{\text{def}}{=} L(r_1) \cup L(r_2)$$

$$L(r_1 \cdot r_2) \stackrel{\text{def}}{=} L(r_1) @ L(r_2)$$

$$L(r^*) \stackrel{\text{def}}{=} \bigcup_{n \geq 0} L(r)^n$$

L is a function from
regular expressions to sets
of strings

$$L : \text{Rexp} \Rightarrow \text{Set}[\text{String}]$$

The Meaning of a Regular Expression

$$L(\emptyset) \stackrel{\text{def}}{=} \emptyset$$

$$L(\epsilon) \stackrel{\text{def}}{=} \{""\}$$

$$L(c) \stackrel{\text{def}}{=} \{ "c" \}$$

$$L(r_1 + r_2) \stackrel{\text{def}}{=} L(r_1) \cup L(r_2)$$

$$L(r_1 \cdot r_2) \stackrel{\text{def}}{=} L(r_1) @ L(r_2)$$

$$L(r^*) \stackrel{\text{def}}{=} \bigcup_{n \geq 0} L(r)^n$$

$$L(r)^\circ \stackrel{\text{def}}{=} \{""\}$$

$$L(r)^{n+1} \stackrel{\text{def}}{=} L(r) @ L(r)^n$$

L is a function from
regular expressions to sets
of strings

$L : \text{Rexp} \Rightarrow \text{Set}[\text{String}]$

What is $L(a^*)$?

Reg Exp Equivalences

$$(a + b) + c \equiv? a + (b + c)$$

$$a + a \equiv? a$$

$$(a \cdot b) \cdot c \equiv? a \cdot (b \cdot c)$$

$$a \cdot a \equiv? a$$

$$\epsilon^* \equiv? \epsilon$$

$$\emptyset^* \equiv? \emptyset$$

$$\forall r. \quad r \cdot \epsilon \equiv? r$$

$$\forall r. \quad r + \epsilon \equiv? r$$

$$\forall r. \quad r + \emptyset \equiv? r$$

$$\forall r. \quad r \cdot \emptyset \equiv? r$$

$$c \cdot (a + b) \equiv? (c \cdot a) + (c \cdot b)$$

$$a^* \equiv? \epsilon + (a \cdot a^*)$$

Reg Exp Equivalences

$$(a + b) + c \equiv? a + (b + c) \quad \text{yes}$$

$$a + a \equiv? a$$

$$(a \cdot b) \cdot c \equiv? a \cdot (b \cdot c)$$

$$a \cdot a \equiv? a$$

$$\epsilon^* \equiv? \epsilon$$

$$\emptyset^* \equiv? \emptyset$$

$$\forall r. \quad r \cdot \epsilon \equiv? r$$

$$\forall r. \quad r + \epsilon \equiv? r$$

$$\forall r. \quad r + \emptyset \equiv? r$$

$$\forall r. \quad r \cdot \emptyset \equiv? r$$

$$c \cdot (a + b) \equiv? (c \cdot a) + (c \cdot b)$$

$$a^* \equiv? \epsilon + (a \cdot a^*)$$

Reg Exp Equivalences

$$(a + b) + c \equiv? a + (b + c) \quad \text{yes}$$

$$a + a \equiv? a \quad \text{yes}$$

$$(a \cdot b) \cdot c \equiv? a \cdot (b \cdot c)$$

$$a \cdot a \equiv? a$$

$$\epsilon^* \equiv? \epsilon$$

$$\emptyset^* \equiv? \emptyset$$

$$\forall r. \quad r \cdot \epsilon \equiv? r$$

$$\forall r. \quad r + \epsilon \equiv? r$$

$$\forall r. \quad r + \emptyset \equiv? r$$

$$\forall r. \quad r \cdot \emptyset \equiv? r$$

$$c \cdot (a + b) \equiv? (c \cdot a) + (c \cdot b)$$

$$a^* \equiv? \epsilon + (a \cdot a^*)$$

Reg Exp Equivalences

$$(a + b) + c \equiv? a + (b + c) \quad \text{yes}$$

$$a + a \equiv? a \quad \text{yes}$$

$$(a \cdot b) \cdot c \equiv? a \cdot (b \cdot c) \quad \text{yes}$$

$$a \cdot a \equiv? a$$

$$\epsilon^* \equiv? \epsilon$$

$$\emptyset^* \equiv? \emptyset$$

$$\forall r. \quad r \cdot \epsilon \equiv? r$$

$$\forall r. \quad r + \epsilon \equiv? r$$

$$\forall r. \quad r + \emptyset \equiv? r$$

$$\forall r. \quad r \cdot \emptyset \equiv? r$$

$$c \cdot (a + b) \equiv? (c \cdot a) + (c \cdot b)$$

$$a^* \equiv? \epsilon + (a \cdot a^*)$$

Reg Exp Equivalences

$$(a + b) + c \equiv? a + (b + c) \quad \text{yes}$$

$$a + a \equiv? a \quad \text{yes}$$

$$(a \cdot b) \cdot c \equiv? a \cdot (b \cdot c) \quad \text{yes}$$

$$a \cdot a \equiv? a \quad \text{no}$$

$$\epsilon^* \equiv? \epsilon$$

$$\emptyset^* \equiv? \emptyset$$

$$\forall r. \quad r \cdot \epsilon \equiv? r$$

$$\forall r. \quad r + \epsilon \equiv? r$$

$$\forall r. \quad r + \emptyset \equiv? r$$

$$\forall r. \quad r \cdot \emptyset \equiv? r$$

$$c \cdot (a + b) \equiv? (c \cdot a) + (c \cdot b)$$

$$a^* \equiv? \epsilon + (a \cdot a^*)$$

Reg Exp Equivalences

$$(a + b) + c \equiv? a + (b + c) \quad \text{yes}$$

$$a + a \equiv? a \quad \text{yes}$$

$$(a \cdot b) \cdot c \equiv? a \cdot (b \cdot c) \quad \text{yes}$$

$$a \cdot a \equiv? a \quad \text{no}$$

$$\epsilon^* \equiv? \epsilon \quad \text{yes}$$

$$\emptyset^* \equiv? \emptyset$$

$$\forall r. \quad r \cdot \epsilon \equiv? r$$

$$\forall r. \quad r + \epsilon \equiv? r$$

$$\forall r. \quad r + \emptyset \equiv? r$$

$$\forall r. \quad r \cdot \emptyset \equiv? r$$

$$c \cdot (a + b) \equiv? (c \cdot a) + (c \cdot b)$$

$$a^* \equiv? \epsilon + (a \cdot a^*)$$

Reg Exp Equivalences

$$(a + b) + c \equiv? a + (b + c) \quad \text{yes}$$

$$a + a \equiv? a \quad \text{yes}$$

$$(a \cdot b) \cdot c \equiv? a \cdot (b \cdot c) \quad \text{yes}$$

$$a \cdot a \equiv? a \quad \text{no}$$

$$\epsilon^* \equiv? \epsilon \quad \text{yes}$$

$$\emptyset^* \equiv? \emptyset \quad \text{no}$$

$$\forall r. \quad r \cdot \epsilon \equiv? r$$

$$\forall r. \quad r + \epsilon \equiv? r$$

$$\forall r. \quad r + \emptyset \equiv? r$$

$$\forall r. \quad r \cdot \emptyset \equiv? r$$

$$c \cdot (a + b) \equiv? (c \cdot a) + (c \cdot b)$$

$$a^* \equiv? \epsilon + (a \cdot a^*)$$

Reg Exp Equivalences

$$(a + b) + c \equiv? a + (b + c) \quad \text{yes}$$

$$a + a \equiv? a \quad \text{yes}$$

$$(a \cdot b) \cdot c \equiv? a \cdot (b \cdot c) \quad \text{yes}$$

$$a \cdot a \equiv? a \quad \text{no}$$

$$\epsilon^* \equiv? \epsilon \quad \text{yes}$$

$$\emptyset^* \equiv? \emptyset \quad \text{no}$$

$$\forall r. \quad r \cdot \epsilon \equiv? r \quad \text{yes}$$

$$\forall r. \quad r + \epsilon \equiv? r$$

$$\forall r. \quad r + \emptyset \equiv? r$$

$$\forall r. \quad r \cdot \emptyset \equiv? r$$

$$c \cdot (a + b) \equiv? (c \cdot a) + (c \cdot b)$$

$$a^* \equiv? \epsilon + (a \cdot a^*)$$

Reg Exp Equivalences

$$(a + b) + c \equiv? a + (b + c) \quad \text{yes}$$

$$a + a \equiv? a \quad \text{yes}$$

$$(a \cdot b) \cdot c \equiv? a \cdot (b \cdot c) \quad \text{yes}$$

$$a \cdot a \equiv? a \quad \text{no}$$

$$\epsilon^* \equiv? \epsilon \quad \text{yes}$$

$$\emptyset^* \equiv? \emptyset \quad \text{no}$$

$$\forall r. \quad r \cdot \epsilon \equiv? r \quad \text{yes}$$

$$\forall r. \quad r + \epsilon \equiv? r \quad \text{no}$$

$$\forall r. \quad r + \emptyset \equiv? r$$

$$\forall r. \quad r \cdot \emptyset \equiv? r$$

$$c \cdot (a + b) \equiv? (c \cdot a) + (c \cdot b)$$

$$a^* \equiv? \epsilon + (a \cdot a^*)$$

Reg Exp Equivalences

	$(a + b) + c$	$\equiv?$	$a + (b + c)$	yes
	$a + a$	$\equiv?$	a	yes
	$(a \cdot b) \cdot c$	$\equiv?$	$a \cdot (b \cdot c)$	yes
	$a \cdot a$	$\equiv?$	a	no
	ϵ^*	$\equiv?$	ϵ	yes
	\emptyset^*	$\equiv?$	\emptyset	no
$\forall r.$	$r \cdot \epsilon$	$\equiv?$	r	yes
$\forall r.$	$r + \epsilon$	$\equiv?$	r	no
$\forall r.$	$r + \emptyset$	$\equiv?$	r	yes
$\forall r.$	$r \cdot \emptyset$	$\equiv?$	r	
	$c \cdot (a + b)$	$\equiv?$	$(c \cdot a) + (c \cdot b)$	
	a^*	$\equiv?$	$\epsilon + (a \cdot a^*)$	

Reg Exp Equivalences

	$(a + b) + c$	$\equiv?$	$a + (b + c)$	yes
	$a + a$	$\equiv?$	a	yes
	$(a \cdot b) \cdot c$	$\equiv?$	$a \cdot (b \cdot c)$	yes
	$a \cdot a$	$\equiv?$	a	no
	ϵ^*	$\equiv?$	ϵ	yes
	\emptyset^*	$\equiv?$	\emptyset	no
$\forall r.$	$r \cdot \epsilon$	$\equiv?$	r	yes
$\forall r.$	$r + \epsilon$	$\equiv?$	r	no
$\forall r.$	$r + \emptyset$	$\equiv?$	r	yes
$\forall r.$	$r \cdot \emptyset$	$\equiv?$	r	no
	$c \cdot (a + b)$	$\equiv?$	$(c \cdot a) + (c \cdot b)$	
	a^*	$\equiv?$	$\epsilon + (a \cdot a^*)$	

Reg Exp Equivalences

	$(a + b) + c$	$\equiv?$	$a + (b + c)$	yes
	$a + a$	$\equiv?$	a	yes
	$(a \cdot b) \cdot c$	$\equiv?$	$a \cdot (b \cdot c)$	yes
	$a \cdot a$	$\equiv?$	a	no
	ϵ^*	$\equiv?$	ϵ	yes
	\emptyset^*	$\equiv?$	\emptyset	no
$\forall r.$	$r \cdot \epsilon$	$\equiv?$	r	yes
$\forall r.$	$r + \epsilon$	$\equiv?$	r	no
$\forall r.$	$r + \emptyset$	$\equiv?$	r	yes
$\forall r.$	$r \cdot \emptyset$	$\equiv?$	r	no
	$c \cdot (a + b)$	$\equiv?$	$(c \cdot a) + (c \cdot b)$	yes
	a^*	$\equiv?$	$\epsilon + (a \cdot a^*)$	

Reg Exp Equivalences

	$(a + b) + c$	$\equiv?$	$a + (b + c)$	yes
	$a + a$	$\equiv?$	a	yes
	$(a \cdot b) \cdot c$	$\equiv?$	$a \cdot (b \cdot c)$	yes
	$a \cdot a$	$\equiv?$	a	no
	ϵ^*	$\equiv?$	ϵ	yes
	\emptyset^*	$\equiv?$	\emptyset	no
$\forall r.$	$r \cdot \epsilon$	$\equiv?$	r	yes
$\forall r.$	$r + \epsilon$	$\equiv?$	r	no
$\forall r.$	$r + \emptyset$	$\equiv?$	r	yes
$\forall r.$	$r \cdot \emptyset$	$\equiv?$	r	no
	$c \cdot (a + b)$	$\equiv?$	$(c \cdot a) + (c \cdot b)$	yes
	a^*	$\equiv?$	$\epsilon + (a \cdot a^*)$	yes

The Specification for Matching

a regular expression r matches a string s
if and only if

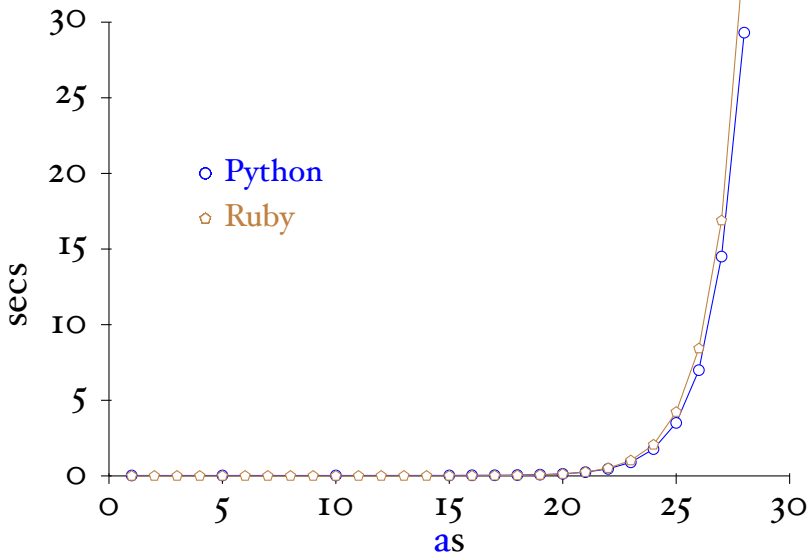
$$s \in L(r)$$

The Specification for Matching

a regular expression r matches a string s
if and only if

$$s \in L(r)$$

$$(a^{\{n\}}) \cdot a^{\{n\}}$$



Evil Regular Expressions

- Regular expression Denial of Service (ReDoS)
- Evil regular expressions
 - $(a?\{n\}) \cdot a\{n\}$
 - $(a^+)^+$
 - $([a-z]^+)^*$
 - $(a + a \cdot a)^+$
 - $(a + a?)^+$

A Matching Algorithm

...whether a regular expression can match the empty string:

$$\text{nullable}(\emptyset) \stackrel{\text{def}}{=} \textit{false}$$

$$\text{nullable}(\epsilon) \stackrel{\text{def}}{=} \textit{true}$$

$$\text{nullable}(c) \stackrel{\text{def}}{=} \textit{false}$$

$$\text{nullable}(r_1 + r_2) \stackrel{\text{def}}{=} \text{nullable}(r_1) \vee \text{nullable}(r_2)$$

$$\text{nullable}(r_1 \cdot r_2) \stackrel{\text{def}}{=} \text{nullable}(r_1) \wedge \text{nullable}(r_2)$$

$$\text{nullable}(r^*) \stackrel{\text{def}}{=} \textit{true}$$

A Matching Algorithm

...whether a regular expression can match the empty string:

$nullable(\emptyset) \stackrel{\text{def}}{=} \text{false}$

$nullable(\epsilon) \stackrel{\text{def}}{=} \text{true}$

$nullable(c) \stackrel{\text{def}}{=} \text{false}$

$nullable(r_1 + r_2) \stackrel{\text{def}}{=} nullable(r_1) \vee nullable(r_2)$

$nullable(r_1 \cdot r_2) \stackrel{\text{def}}{=} nullable(r_1) \wedge nullable(r_2)$

$nullable(n)$

```
def nullable (r: Rexp) : Boolean = r match {  
  case NULL => false  
  case EMPTY => true  
  case CHAR(_) => false  
  case ALT(r1, r2) => nullable(r1) || nullable(r2)  
  case SEQ(r1, r2) => nullable(r1) && nullable(r2)  
  case STAR(_) => true  
}
```

The Derivative of a Rexp

If r matches the string $c::s$, what is a regular expression that matches s ?

$der\ c\ r$ gives the answer

The Derivative of a Rexp (2)

$$\mathit{der} c (\emptyset) \stackrel{\text{def}}{=} \emptyset$$

$$\mathit{der} c (\epsilon) \stackrel{\text{def}}{=} \emptyset$$

$$\mathit{der} c (d) \stackrel{\text{def}}{=} \text{if } c = d \text{ then } \epsilon \text{ else } \emptyset$$

$$\mathit{der} c (r_1 + r_2) \stackrel{\text{def}}{=} \mathit{der} c r_1 + \mathit{der} c r_2$$

$$\mathit{der} c (r_1 \cdot r_2) \stackrel{\text{def}}{=} \text{if } \mathit{nullable}(r_1) \\ \text{then } (\mathit{der} c r_1) \cdot r_2 + \mathit{der} c r_2 \\ \text{else } (\mathit{der} c r_1) \cdot r_2$$

$$\mathit{der} c (r^*) \stackrel{\text{def}}{=} (\mathit{der} c r) \cdot (r^*)$$

The Derivative of a Rexp (2)

$$\mathit{der} c (\emptyset) \stackrel{\text{def}}{=} \emptyset$$

$$\mathit{der} c (\epsilon) \stackrel{\text{def}}{=} \emptyset$$

$$\mathit{der} c (d) \stackrel{\text{def}}{=} \text{if } c = d \text{ then } \epsilon \text{ else } \emptyset$$

$$\mathit{der} c (r_1 + r_2) \stackrel{\text{def}}{=} \mathit{der} c r_1 + \mathit{der} c r_2$$

$$\mathit{der} c (r_1 \cdot r_2) \stackrel{\text{def}}{=} \text{if } \mathit{nullable}(r_1) \\ \text{then } (\mathit{der} c r_1) \cdot r_2 + \mathit{der} c r_2 \\ \text{else } (\mathit{der} c r_1) \cdot r_2$$

$$\mathit{der} c (r^*) \stackrel{\text{def}}{=} (\mathit{der} c r) \cdot (r^*)$$

$$\mathit{ders} [] r \stackrel{\text{def}}{=} r$$

$$\mathit{ders} (c :: s) r \stackrel{\text{def}}{=} \mathit{ders} s (\mathit{der} c r)$$

The Derivative of a Rexp (2)

$der\ c(\emptyset) \stackrel{\text{def}}{=} \emptyset$

$der\ c(\epsilon) \stackrel{\text{def}}{=} \emptyset$

```
def der (r: Rexp, c: Char) : Rexp = r match {
  case NULL => NULL
  case EMPTY => NULL
  case CHAR(d) => if (c == d) EMPTY else NULL
  case ALT(r1, r2) => ALT(der(r1, c), der(r2, c))
  case SEQ(r1, r2) =>
    if (nullable(r1)) ALT(SEQ(der(r1, c), r2), der(r2, c))
    else SEQ(der(r1, c), r2)
  case STAR(r) => SEQ(der(r, c), STAR(r))
}
```



```
def ders (s: List[Char], r: Rexp) : Rexp = s match {
  case Nil => r
  case c::s => ders(s, der(c, r))
}
```

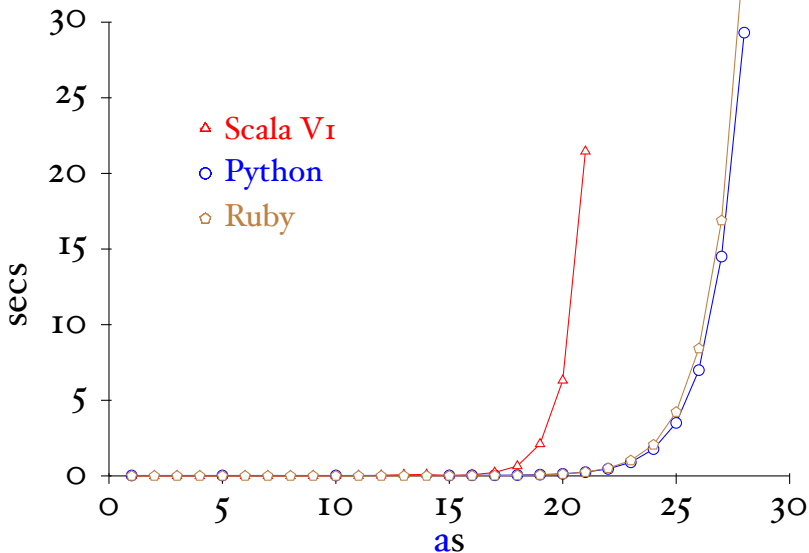

Examples

Given $r \stackrel{\text{def}}{=} ((a \cdot b) + b)^*$ what is

der a r

der b r

$$(a?\{n\}) \cdot a\{n\}$$



Proofs about Rexps

Remember their inductive definition:

$$r ::= \begin{array}{l} \emptyset \\ \epsilon \\ c \\ r_1 \cdot r_2 \\ r_1 + r_2 \\ r^* \end{array}$$

If we want to prove something, say a property $P(r)$, for all regular expressions r then ...

Proofs about Rexp (2)

- P holds for \emptyset , ϵ and c
- P holds for $r_1 + r_2$ under the assumption that P already holds for r_1 and r_2 .
- P holds for $r_1 \cdot r_2$ under the assumption that P already holds for r_1 and r_2 .
- P holds for r^* under the assumption that P already holds for r .

Proofs about Rexp (3)

Assume $P(r)$ is the property:

nullable(r) if and only if $\epsilon \in L(r)$

Proofs about Rexp (4)

Let $Der\ c\ A$ be the set defined as

$$Der\ c\ A \stackrel{\text{def}}{=} \{s \mid c :: s \in A\}$$

We can prove

$$L(\text{der}\ c\ r) = Der\ c\ (L(r))$$

by induction on r .

Proofs about Strings

If we want to prove something, say a property $P(s)$, for all strings s then ...

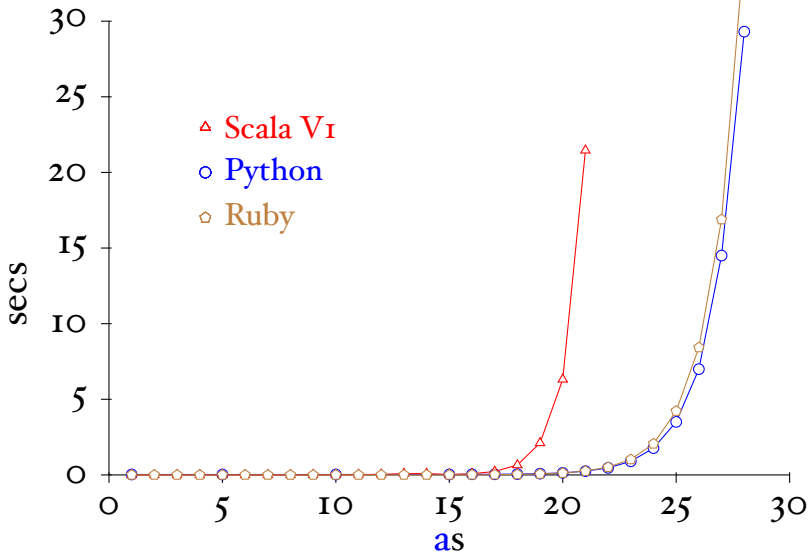
- P holds for the empty string, and
- P holds for the string $c::s$ under the assumption that P already holds for s

Proofs about Strings (2)

We can finally prove

matcher(r, s) if and only if $s \in L(r)$

$$(a?\{n\}) \cdot a\{n\}$$



A Problem

We represented the “n-times” $a\{n\}$ as a sequence regular expression:

1: a

2: $a \cdot a$

3: $a \cdot a \cdot a$

...

13: $a \cdot a \cdot a \cdot a \cdot a \cdot a \cdot a \cdot a \cdot a \cdot a \cdot a \cdot a \cdot a$

...

20:

This problem is aggravated with $a?$ being represented as $\epsilon + a$.

Solving the Problem

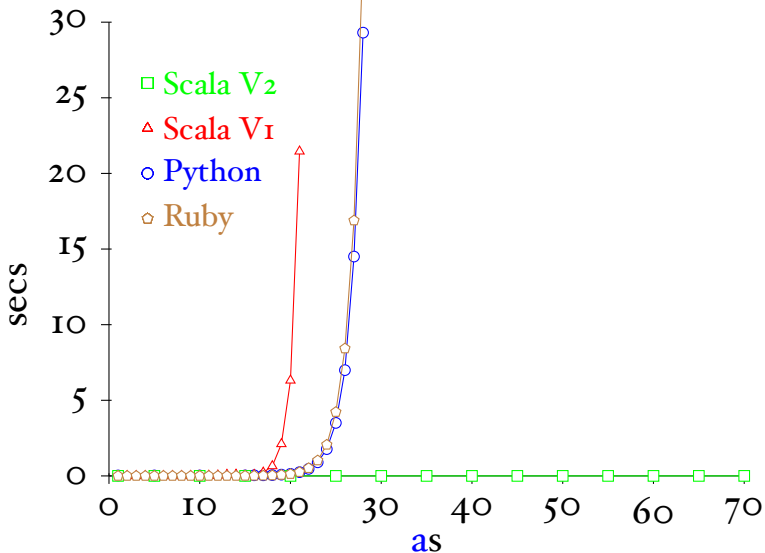
What happens if we extend our regular expressions

$$r ::= \dots$$

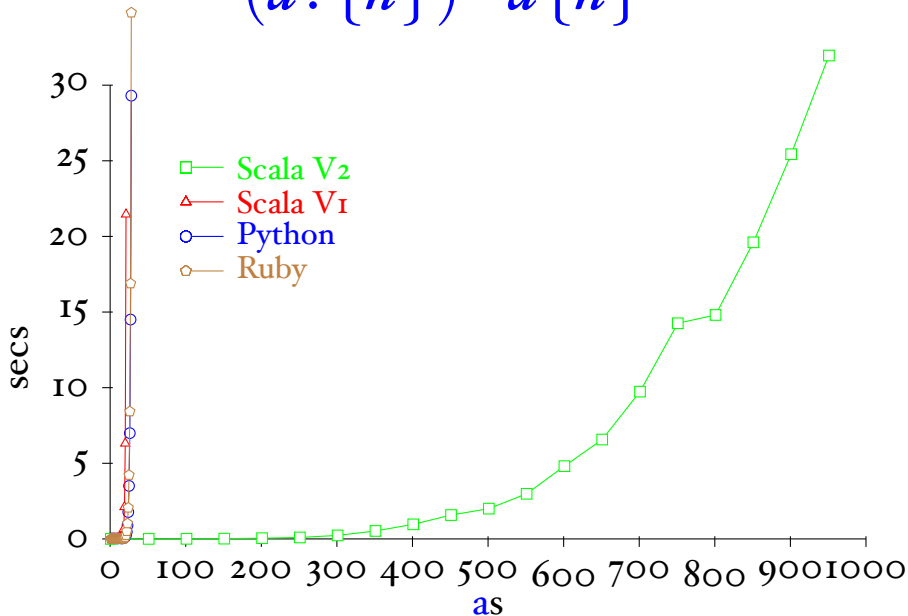
		$r\{n\}$
		$r?$

What is their meaning? What are the cases for *nullable* and *der*?

$$(a?\{n\}) \cdot a\{n\}$$



$$(a^{? \{n\}}) \cdot a\{n\}$$



Examples

Recall the example of $r \stackrel{\text{def}}{=} ((a \cdot b) + b)^*$ with

$$\text{der } a r = ((\epsilon \cdot b) + \emptyset) \cdot r$$

$$\text{der } b r = ((\emptyset \cdot b) + \epsilon) \cdot r$$

What are these regular expressions equal to?

$$(a?\{n\}) \cdot a\{n\}$$

