UNIVERSITY OF CAMBRIDGE

Weighted finite-state automata

Rogier van Dalen

19 April 2012

Cambridge University Engineering Department

Outline

Finite-state automata

- Motivating example: a spell checker
- The language model: a weighted automaton
- The spelling model: a transducer
- The error model: a weighted transducer

Parameter estimation

Rational kernels

Outline

Finite-state acceptor

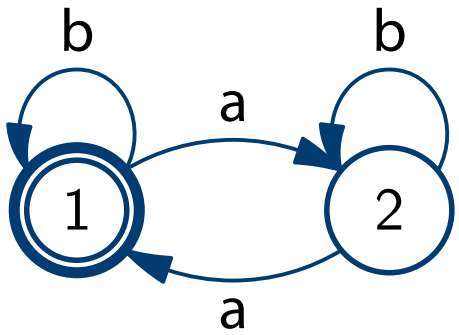
A finite-state acceptor accepts sequences of symbols that correspond to paths from a start state to a final state.

State (label has no meaning)

Start state

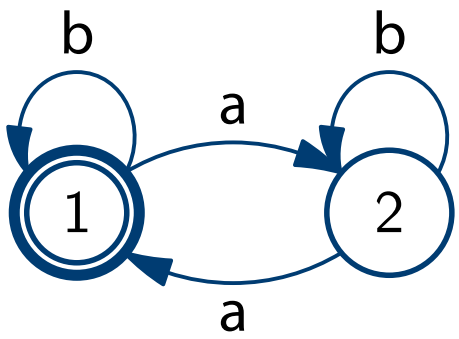
Final state

Transition that accepts symbol x



Finite-state automata

This is not a graphical model. It is blue.

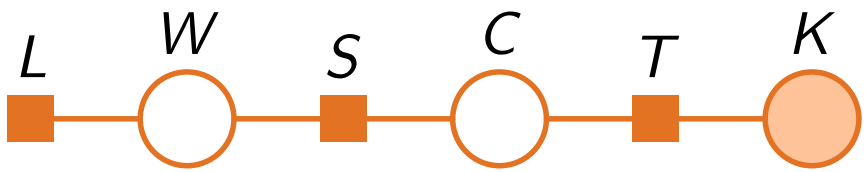


We will see

- weighted automata;
- n-tape automata.
  - $n = 2$: transducers.

Motivating example: a spell checker

This is a graphical model. It is orange and has factors.



with

- W the words;
- C the correct characters;
- K the key presses;

and

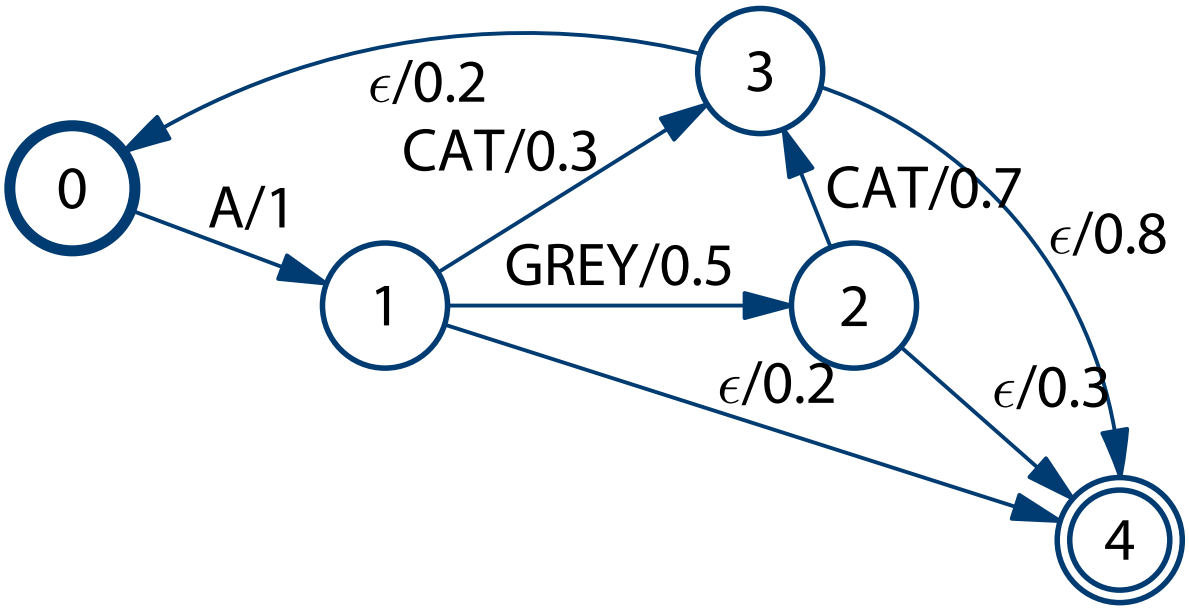
- L the language model;
- S the spelling model;
- T the error (typo) model.

The language model L: a weighted automaton

n-gram model:

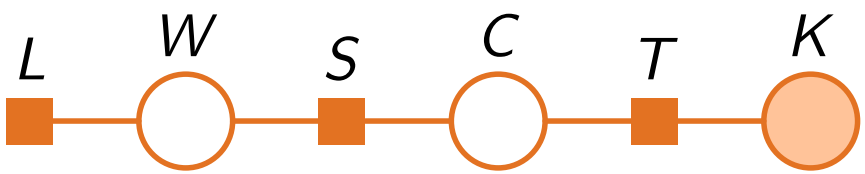
$$P(W) \simeq P(w_1)\,P(w_2|w_1)\,P(w_3|w_2)\cdots P(w_n|w_{n-1})$$
$$\triangleq [[L]](W)$$

$L =$

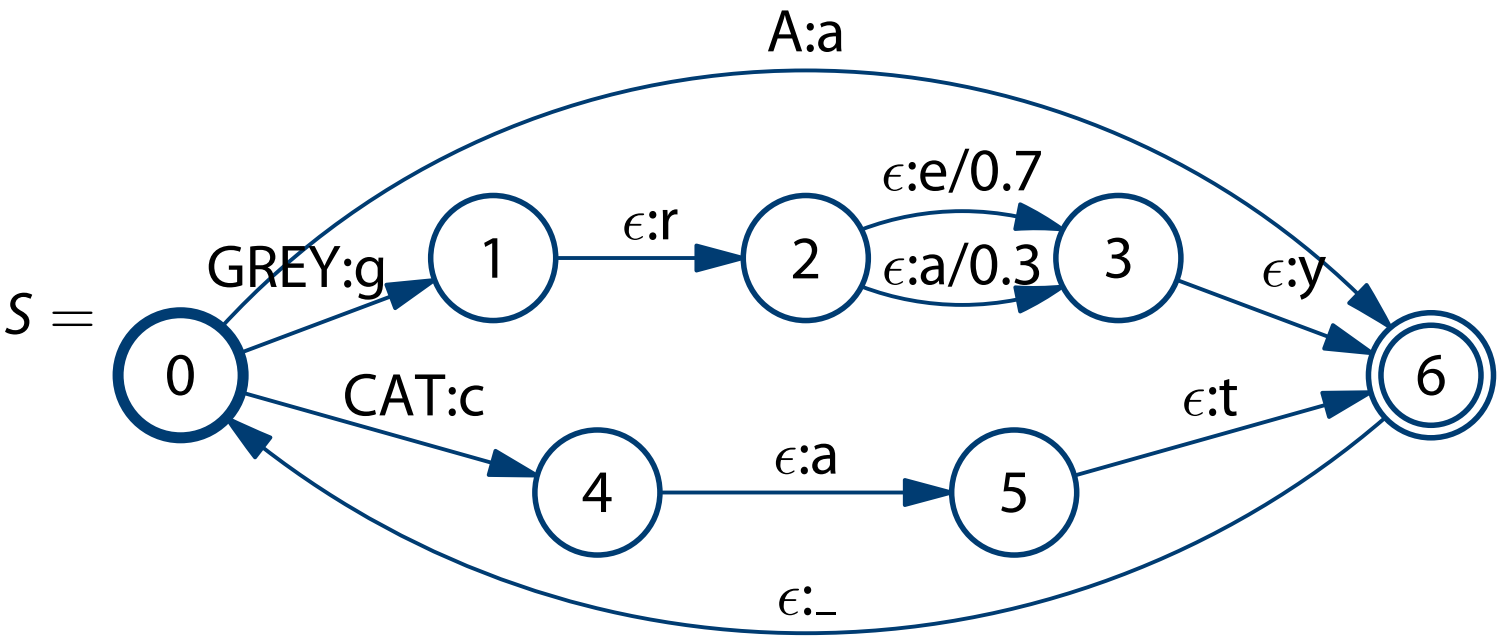


- /0.3 indicates a weight of 0.3.
- ϵ means no symbol.

The spelling model: a transducer

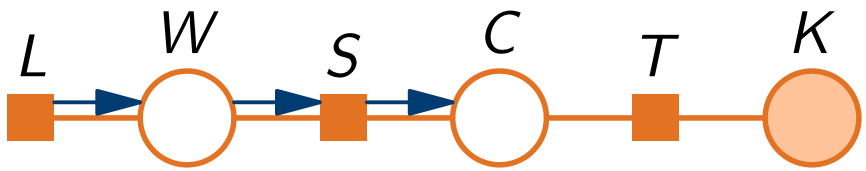


$$P(C|W) = [[S]](W, C)$$



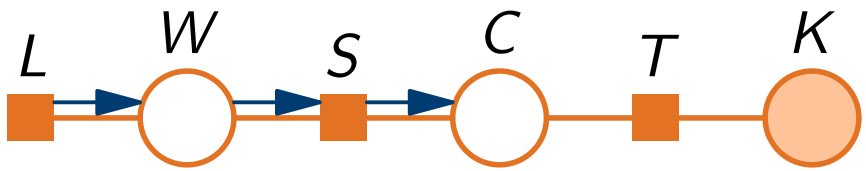
- a:b indicates transduction from a to b.

Message passing



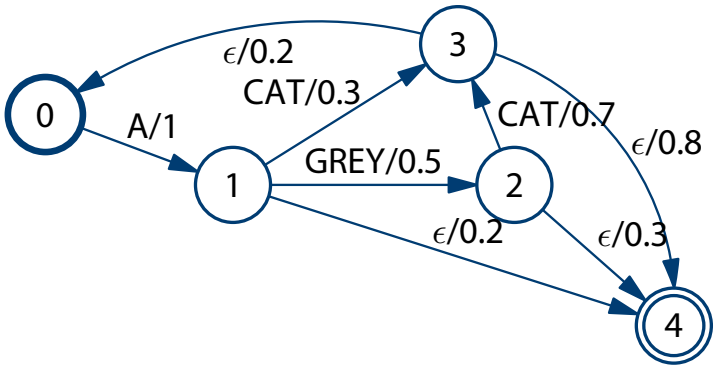
What are the messages?

Message passing

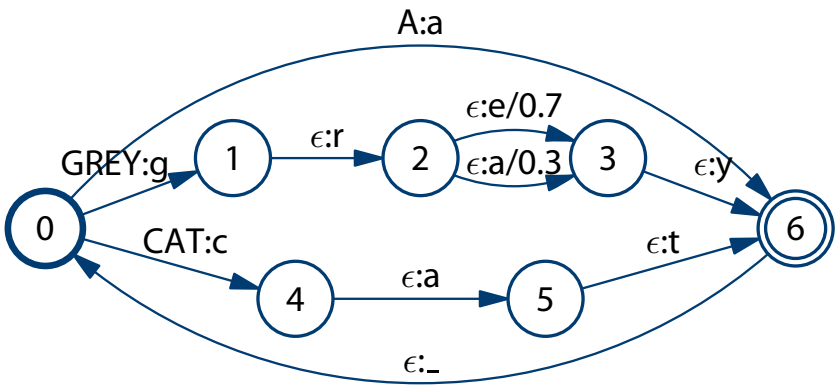


What are the messages?

$L =$

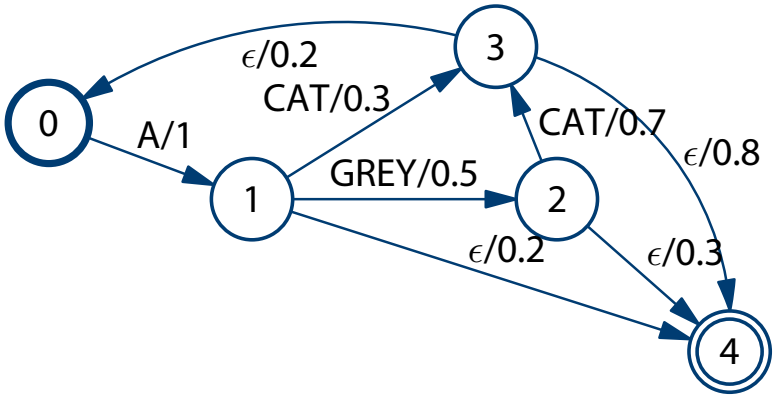


$S =$

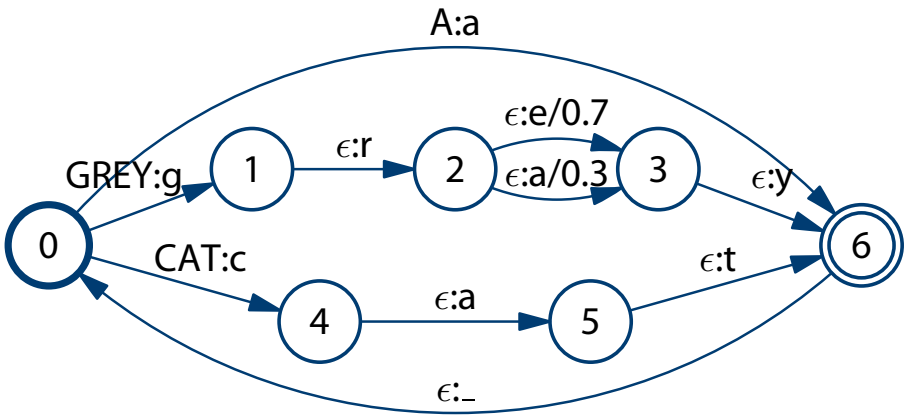


Composition

$L =$

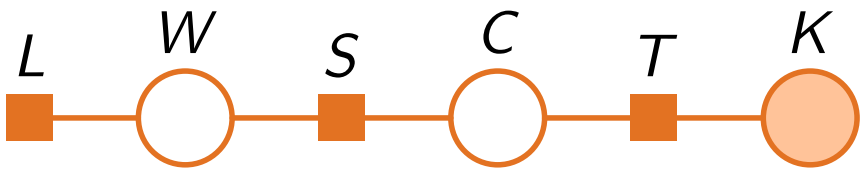


$S =$



- Composition: $L \circ S$;
- projection.

The error model: a weighted transducer

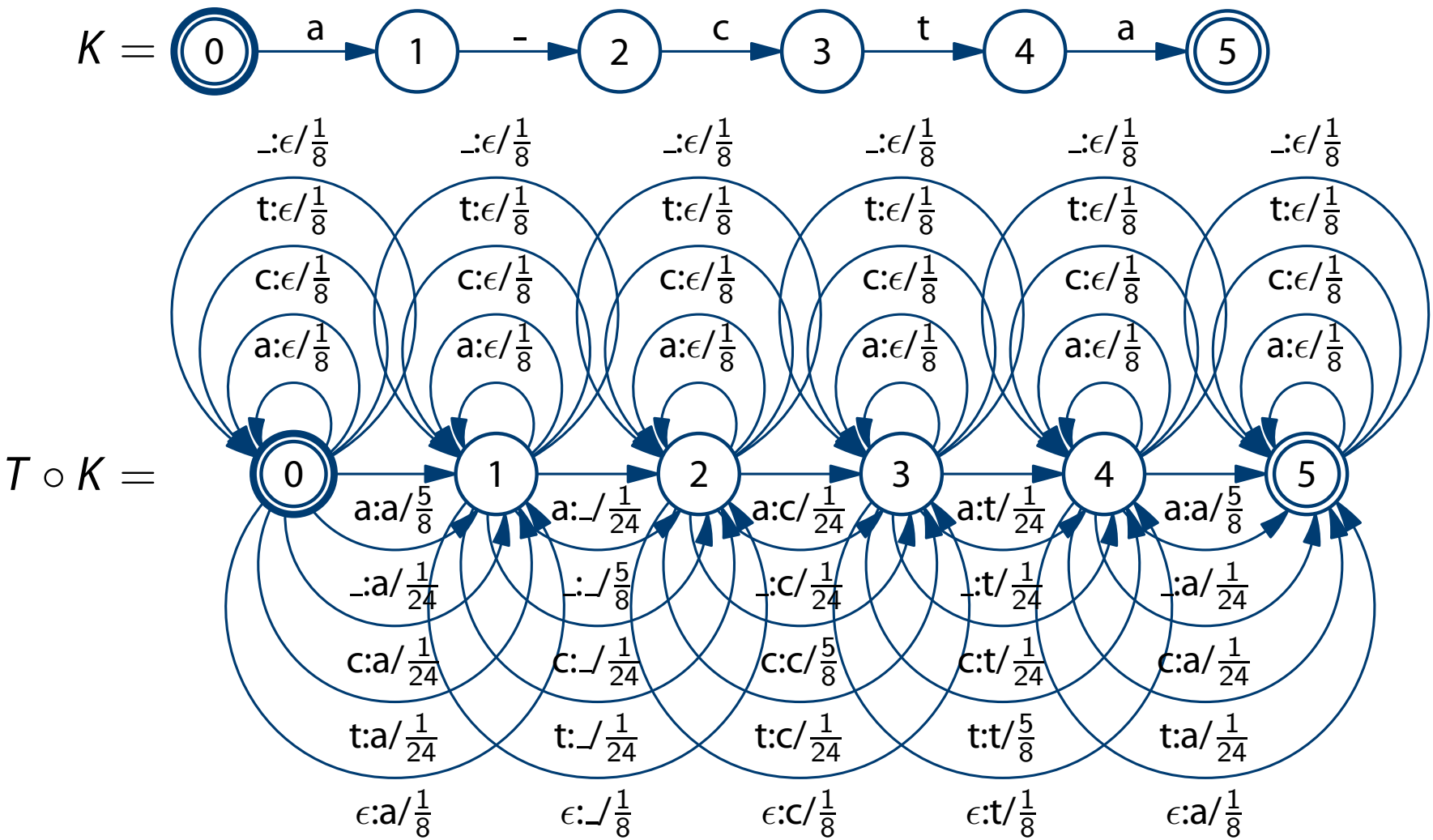


$$P(K|C) \propto [[T]](C,K)$$

$$T = \epsilon{:}\_/\tfrac{1}{32} \;\; \epsilon{:}t/\tfrac{1}{32} \;\; \epsilon{:}c/\tfrac{1}{32} \;\; \epsilon{:}a/\tfrac{1}{32} \;\; 0 \;\; a{:}\epsilon/\tfrac{1}{8} \;\; c{:}\epsilon/\tfrac{1}{8} \;\; t{:}\epsilon/\tfrac{1}{8} \;\; \_{:}\epsilon/\tfrac{1}{8}$$

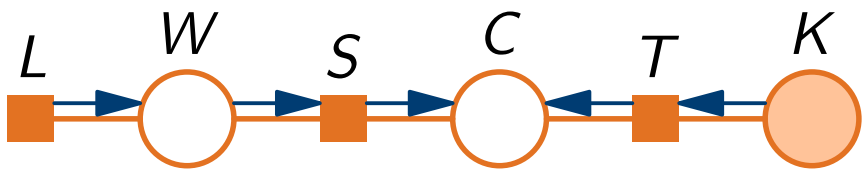
$\_{:}\_/\frac{5}{8}$
$t{:}t/\frac{5}{8}$
$c{:}c/\frac{5}{8}$
$a{:}a/\frac{5}{8}$
$a{:}c/\frac{1}{24}$
$a{:}t/\frac{1}{24}$
$a{:}\_\epsilon/\frac{1}{24}$
$c{:}a{:}\epsilon/\frac{1}{8}$
...

Composition



What does it mean to have multiple paths with input "a _ c a t"?

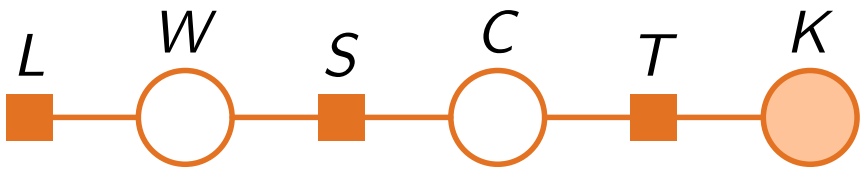
Different semirings for weights



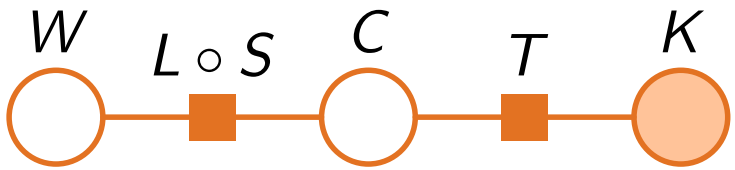
- $L \circ S \circ T \circ K$ scales linearly with the input.
- $\max_C [[L \circ S \circ T \circ K]] (C)$ means summing over all paths.
- Instead, take the best path.
- Formally, use a different semiring for the weights:
  - $(\mathbb{K}, \oplus, \otimes, \overline{0}, \overline{1}) = (\mathbb{R}, +, \times, 0, 1)$;
  - $(\mathbb{K}, \oplus, \otimes, \overline{0}, \overline{1}) = (\mathbb{R}, \max, \times, 0, 1)$;
  - and their logarithms (e.g. "tropical semiring").
- Always use a "shortest-distance" algorithm.
- Roughly equivalent to forms of message-passing:
  - sum-product algorithm;
  - max-product algorithm;
  - max-sum algorithm.

Optimisation

- L and S are known beforehand.



- So they can be composed off-line:



- $L \circ S$ can be optimised.
  - Determinisation/minimisation need to know what $\oplus$ means.

Outline

Parameter estimation

- If our parameters are the weights of an automaton, how to estimate them?
- "Expectation semiring" (Eisner 2002).

Outline

Rational kernels

A kernel K is rational (Cortes 2004) if it can be defined as

$$K(x, y) = \psi([[T]](x, y))$$

where

- T is a transducer over some semiring $\mathbb{K}$; and
- $\psi : \mathbb{K} \to \mathbb{R}$.

Positive definite and symmetric kernels

If

- T is a sensible transducer over some semiring $\mathbb{K}$; and
- $\psi : \mathbb{K} \to \mathbb{R}$ is a semiring morphism,

then

$$K(x, y) = \psi([[T \circ T^{-1}]](x, y))$$

is positive definite and symmetric.

Conclusion

Weighted finite-state automata:

- a useful formalism for models over symbol sequences.
- For inference.
- For string kernels.