



# RECONHECIMENTO DE PADRÕES EM TEXTO USANDO AUTÔMATOS FINITOS

Discentes:

Katarina Ires de Castro Pinheiro

Valvan Alves Freitas

Disciplina: Linguagens Formais e Autômatos

# Sumário

- 03 O que são Autômatos Finitos?**
- 04 Elementos dos Autômatos Finitos**
- 05 Aplicações em Reconhecimento de Padrões**
- 06 Pré-processamento de Dados**
- 07 Construção dos Autômatos**
- 08 Padrões em Linguagem Natural (NLP)**
- 09 Métricas de Avaliação**
- 12 Código**
- 13 Resultados e Avaliação**
- 14 Conclusão**

# O que são Autômatos Finitos?

## Definição Formal

Um autômato finito é um modelo matemático que representa um sistema de estados finitos com transições entre eles, baseado em um alfabeto de entrada.

## Elementos Básicos

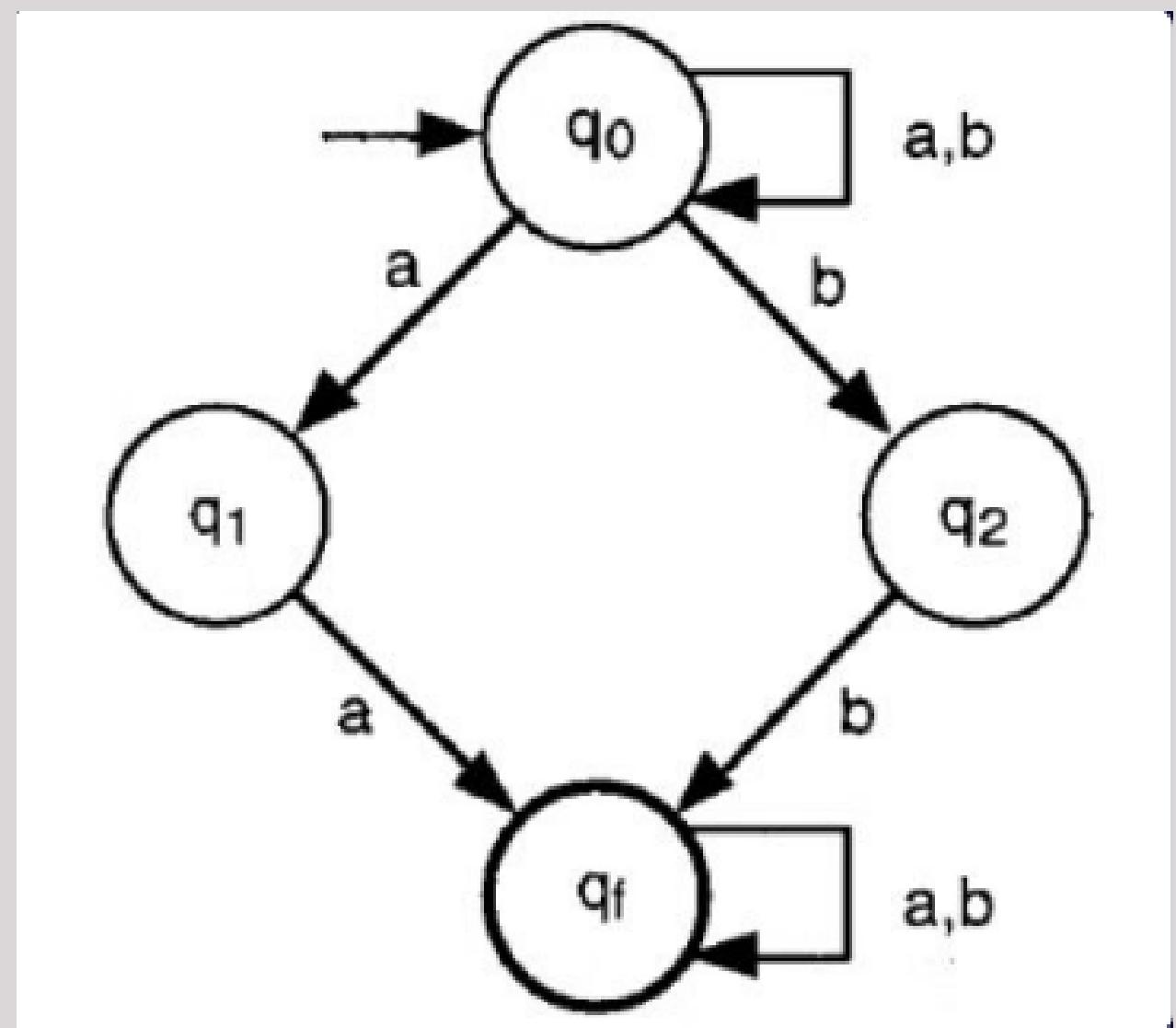
Estados, transições, alfabeto, estado inicial e estados de aceitação são os principais componentes de um autômato finito.

## Aplicabilidade

Os autômatos finitos são amplamente utilizados em diversas áreas da computação, como processamento de linguagem natural e análise de redes.

# Elementos dos Autômatos Finitos

- **Transições:** indicam as mudanças de estado que o autômato pode fazer quando recebe determinadas entradas do alfabeto. Cada transição é definida por um estado de origem, um estado de destino e um símbolo de entrada.
- **Alfabeto:** o alfabeto define os símbolos que podem ser lidos pelo autômato. Geralmente, o alfabeto é composto por letras, números ou outros caracteres que formam a linguagem reconhecida pelo autômato
- **Estado inicial:** representa o estado em que o autômato se encontra antes de realizar qualquer transição
- **Estados de aceitação:** estados finais representam os estados de aceitação, que indicam se uma determinada sequência de entrada é aceita pelo autômato.



# Aplicações em Reconhecimento de Padrões

## RECONHECIMENTO DE TEXTO

Os autômatos finitos são usados para reconhecer padrões em textos, como a identificação de palavras-chave ou a análise de sentimentos em avaliações.

## RECONHECIMENTO DE CARACTERES

Em sistemas de Reconhecimento Óptico de Caracteres (OCR), os autômatos finitos são empregados para identificar e classificar caracteres em imagens digitalizadas.

## APLICAÇÕES FUTURAS

Com o avanço da inteligência artificial, os autômatos finitos têm o potencial de serem aplicados em tarefas cada vez mais complexas de reconhecimento de padrões.

# Pré-processamento de Dados

## Limpeza de Dados

A preparação dos dados de entrada, incluindo a remoção de pontuações e stop words, é essencial para o reconhecimento de padrões eficiente.

## Tokenização

A divisão do texto em tokens (palavras e frases) facilita a identificação de padrões específicos.

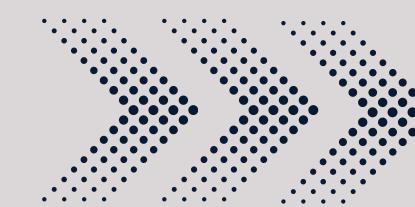
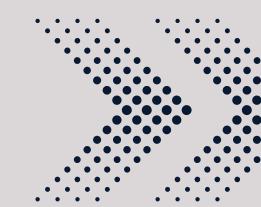
## Normalização

A conversão de todo o texto para minúsculas ou a eliminação de acentuação pode melhorar a precisão do reconhecimento.

## Enriquecimento

A adição de informações contextuais, como metadados, pode enriquecer o processo de reconhecimento de padrões.

# Construção dos Autômatos



## 1º DEFINIÇÃO DE ESTADOS

Os estados do autômato representam as diferentes etapas do reconhecimento de padrões.

## 2º ESPECIFICAÇÃO DE TRANSIÇÕES

As transições entre os estados são definidas com base no alfabeto de entrada, determinando o fluxo do reconhecimento.

## 3º IDENTIFICAÇÃO DE ACEITAÇÃO

Os estados de aceitação indicam quando um padrão específico foi reconhecido com sucesso.

# Processamento de Linguagem Natural (NLP)



## ALIMENTAÇÃO DE ENTRADA

O autômato recebe uma sequência de entradas, como palavras ou caracteres, para processar

## TRANSIÇÃO DE ESTADOS

O autômato navega pelos estados, seguindo as transições definidas, até chegar a um estado de aceitação.

## RECONHECIMENTO DE PADRÃO

Quando o autômato alcança um estado de aceitação, o padrão é considerado reconhecido com sucesso.

# Métricas de avaliação

## PRECISÃO

É a proporção de verdadeiros positivos em relação ao total de previsões positivas feitas pelo modelo.

$$\text{Precision} = \frac{TP}{TP + FP}$$

## RECALL

Também conhecido como sensibilidade, é a proporção de verdadeiros positivos em relação ao total de positivos no conjunto de dados. É útil para avaliar a capacidade do modelo de identificar todas as instâncias relevantes

$$\text{Recall} = \frac{TP}{TP + FN}$$

# Métricas de avaliação

## ACURÁCIA

É a proporção de previsões corretas feitas pelo modelo em relação ao total de previsões feitas. Não recomendada para base de dados desbalanceados.

$$\text{acurácia} = \frac{VP + VN}{VP + FN + VN + FP}$$

## F1-SCORE

É uma métrica que combina precisão e recall em um único valor, calculado a partir da média harmônica dessas duas métricas. É útil quando há desequilíbrio entre as classes do conjunto de dados.

$$\mathbf{F1\ Score} = \frac{2}{\left( \frac{1}{\text{Precision}} + \frac{1}{\text{Recall}} \right)}$$

$$\mathbf{F1\ Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

# Métricas de avaliação

## IoU

Representa a proporção da área de sobreposição entre a predição e o rótulo verdadeiro em relação à área de união dessas duas regiões.

$$IoU = \frac{TP}{(TP + FP + FN)}$$

# Código

# Resultados e Avaliação

- Avaliar o desempenho do reconhecimento de padrões usando autômatos finitos.
- Métricas como precisão, recall e F1-score oferecem uma compreensão completa da eficácia do sistema.
- Essas métricas são calculadas comparando os resultados com um conjunto de dados de referência.
- Os resultado obtidos se mostram condizentes com a proposta do algoritmo, onde foi possível classificar de uma a cinco estrelas o restaurante, baseado nos comentários feitos nos reviews.

# Conclusão

A integração de autômatos finitos com técnicas de Processamento de Linguagem Natural (NLP) viabiliza a construção de modelos sofisticados para a classificação e análise de sentimentos em textos, como demonstrado na predição da satisfação em avaliações. Assim, os autômatos finitos desempenham um papel crucial na evolução da análise de dados textuais, proporcionando avanços significativos em áreas como análise de sentimento e compreensão de linguagem natural.



# OBRIGADO!