



ÁRVORE PATRÍCIA

Arthur, Sérgio e Vitória

O que é?

PATRICIA (Practical Algorithm to Retrieve Information Coded in Alphanumeric) é um algoritmo de busca em árvores, originado das árvores RADIX (Árvore de Prefixos) e Tries.

Basicamente, consiste em uma árvore de caracteres em que seus caminhos internos que possuem apenas um filho serão compactados em apenas um nó contendo os caracteres sucessores da palavra formada a partir do seu nó pai.

História

Surgiu através da motivação de otimizar a busca de títulos em bibliotecas, e em seu uso mais atual, aplica-se nos Sistemas Gerenciadores de Banco de Dados (SGBD), e atua no ganho de eficiência em arquivos XML, além disso, originou a tecnologia utilizada na Blockchain.

A primeira aparição desse termo veio em 1968, cunhado por Donald R. Morrison.

A busca anterior era baseada em índices e catálogos que eram guardados no ambiente físico e ocupavam espaço. Cada título era identificado por uma chave única, e a cada adição de novas obras no acervo, era necessário incluir uma nova chave ou alterar uma já existente, o que limitava muito a capacidade operacional da biblioteca.

História

Assim, foi projetado o algoritmo PATRICIA para as bibliotecas que possuíam computadores, implementarem essa construção de busca para achar os alvos através de suas determinadas chaves usando o índice e o modificando de acordo com novas inserções ou alterações, refletindo essa alteração no sistema e consequentemente na biblioteca.

A solução chegou quando a capacidade de processamento das máquinas, como a memória e processadores, eram mínimas e demoravam muito tempo para obter um retorno. Contudo, o algoritmo em sua forma original foi capaz de solucionar o problema proposto quando ele parecia ser impossível, cumprindo a busca ser feita de forma rápida e automatizando o processo.

Características

Knuth (1973) deu um novo tratamento ao algoritmo, apresentando-o de forma mais clara como um caso particular de pesquisa digital, essencialmente, um caso de árvore trie binária.

Sedgewick(1988) apresentou novos algoritmos de pesquisa e de inserção baseados nos algoritmos propostos por Knuth.

Gonnet e Baeza-Yates (1991) propuseram também outros algoritmos.

Características

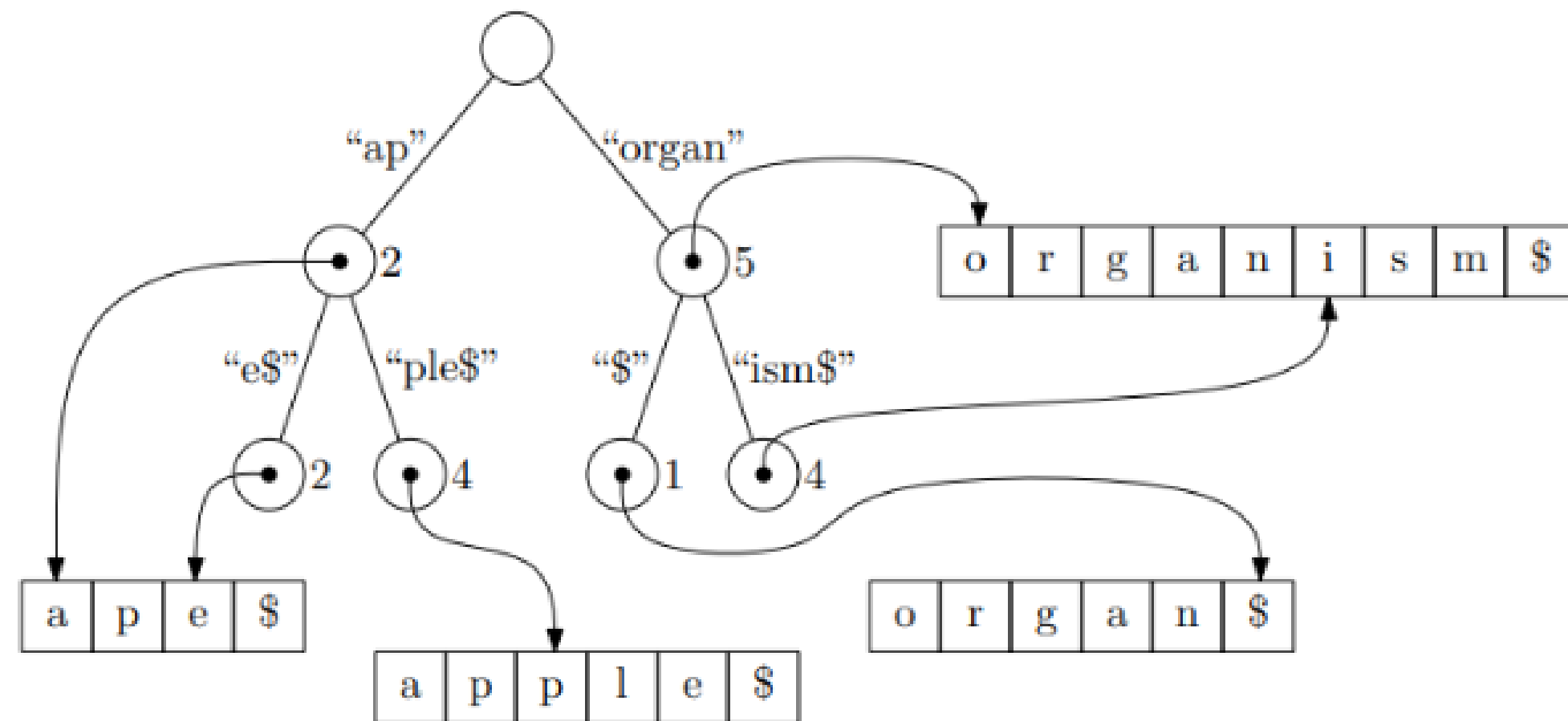
Além disso, a árvore PATRICIA também suporta prefix-matches (ou correspondência por prefixos), nos quais podem retornar uma lista de todas as strings que possuem uma terminação não nula para a string s como prefixo, ou seja, retornam todas as ocorrências de palavras que possuem como prefixo a string s .

Em termos de implementação, é feito através de uma busca comum, verificando todos os caracteres em s até o final.

Características

Se o interesse é encontrar apenas uma string com o prefixo S, é possível verificar o começo da string do último nó verificando no caminho até lá a busca por S.

Esse nó é representado através da representação utilizando altura e ponteiro e esse ponteiro aponta para uma string completa que possui S como prefixo.



Busca

A busca na árvore PATRICIA é similar a busca na Trie, exceto pelo fato da Trie comparar substrings inteiras.

Assim, a busca é feita através de cada nó comparando apenas as substrings, o que otimiza muito mais a busca, sem a necessidade de comparações desnecessárias, dada a configuração dos nós.

Inserção

Inserir uma string em uma árvore PATRICIA é similar a realizar buscas em que a pesquisa chega até ao final, pois a string s não está na árvore.

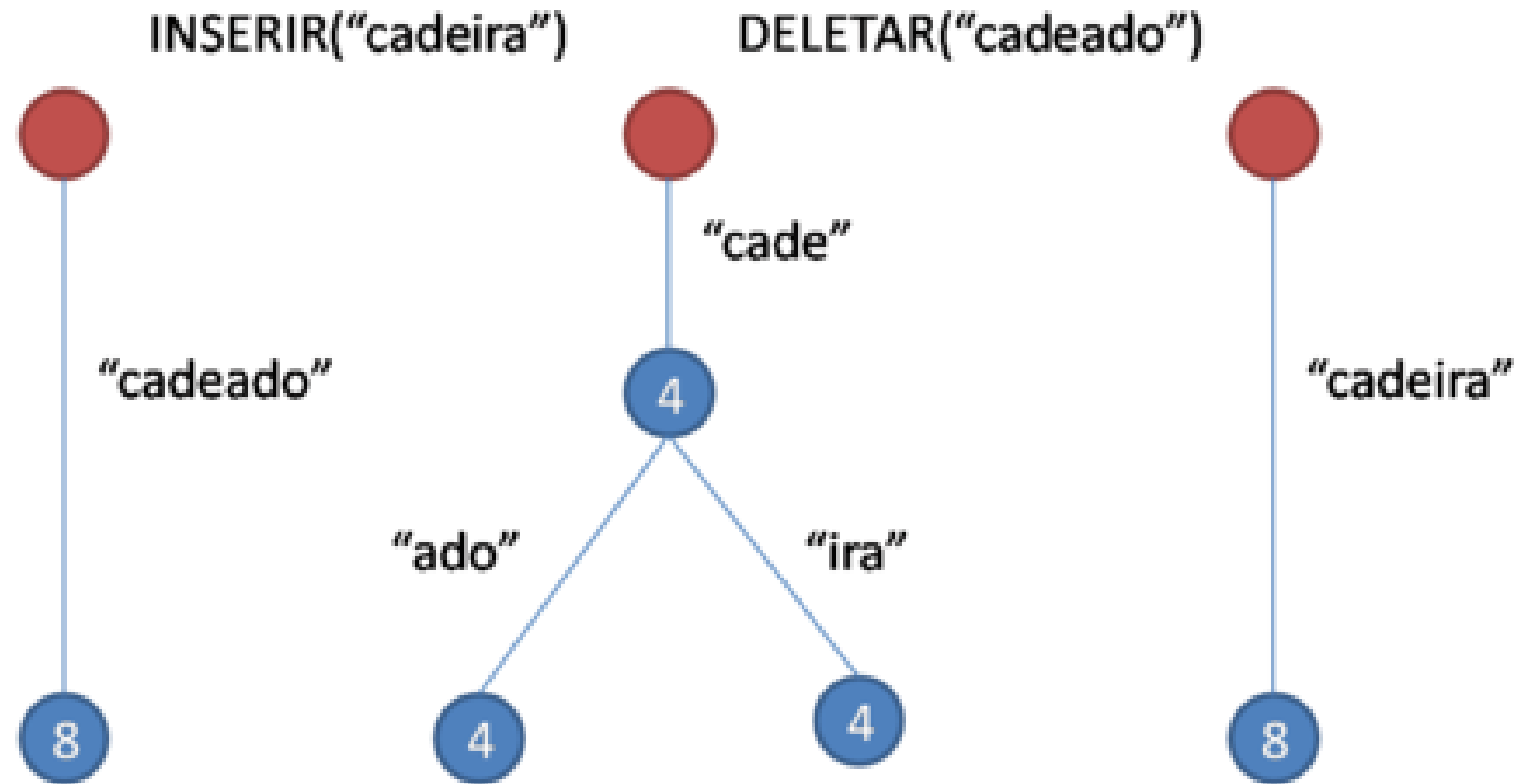
Se a busca é travada no meio de uma aresta A , então A se divide e dá origem a duas novas arestas unidas por um novo nó N .

O remanescente da string s se torna rótulo da aresta, que vai de N até uma folha nova. No entanto, se a pesquisa travar no nó N , então um nó filho é criado e o restante da nova chave é usado como rótulo para aresta entre os dois.

Remoção

A remoção é o oposto da inserção. Primeiramente, é necessário localizar a folha correspondente à string s e removê-la da árvore. Se os parentes u dessa folha estão localizados à esquerda com apenas um filho w , então u é removido e substituído por uma única aresta: unindo o pai de u a w .

Exemplos de Inserção e Remoção



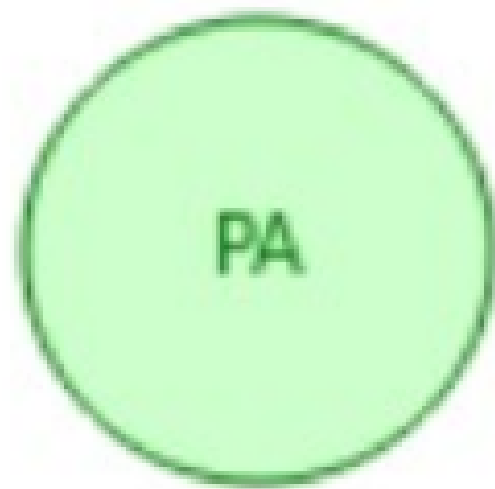
Construção da Árvore

Considere que queremos construir uma árvore Patrícia com os seguintes elementos:

- PATA;
- PASTA;
- PASSA;
- PAPA.

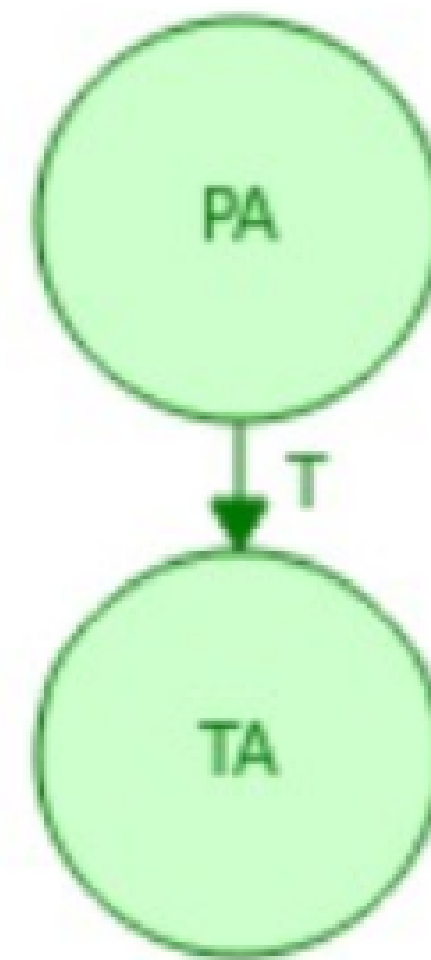
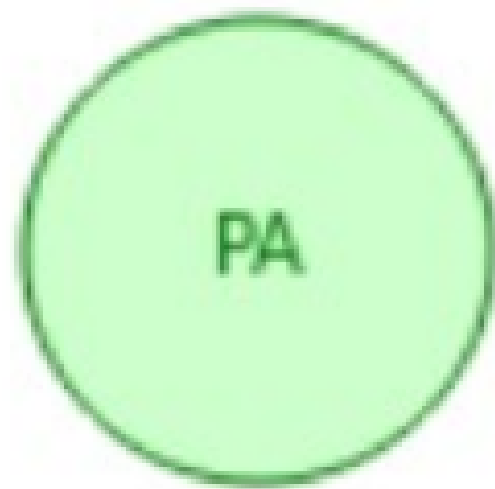
Construção da Árvore

- PA;



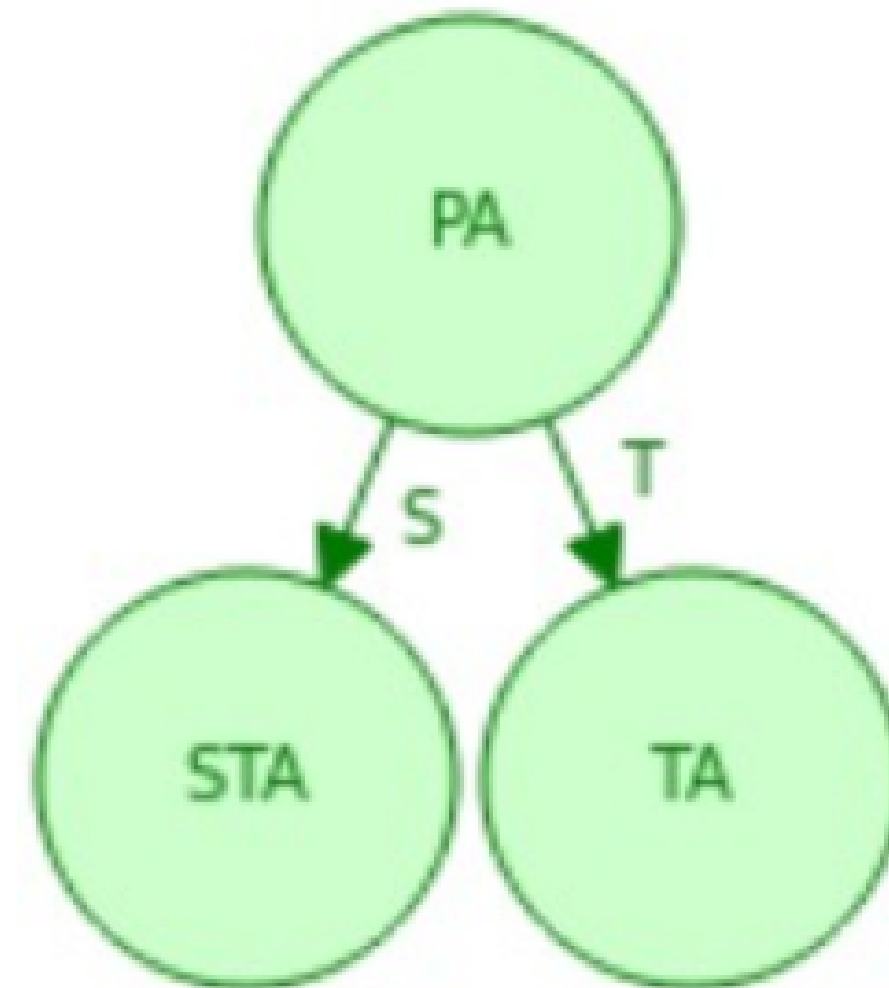
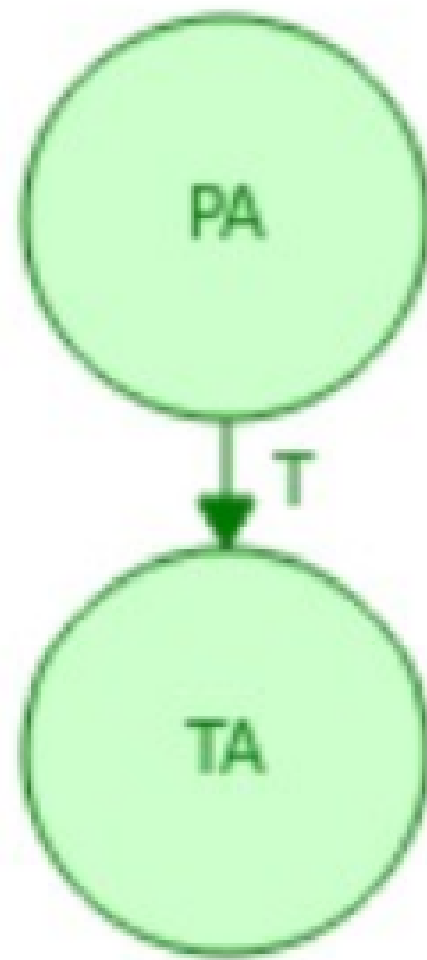
Construção da Árvore

- PATA;



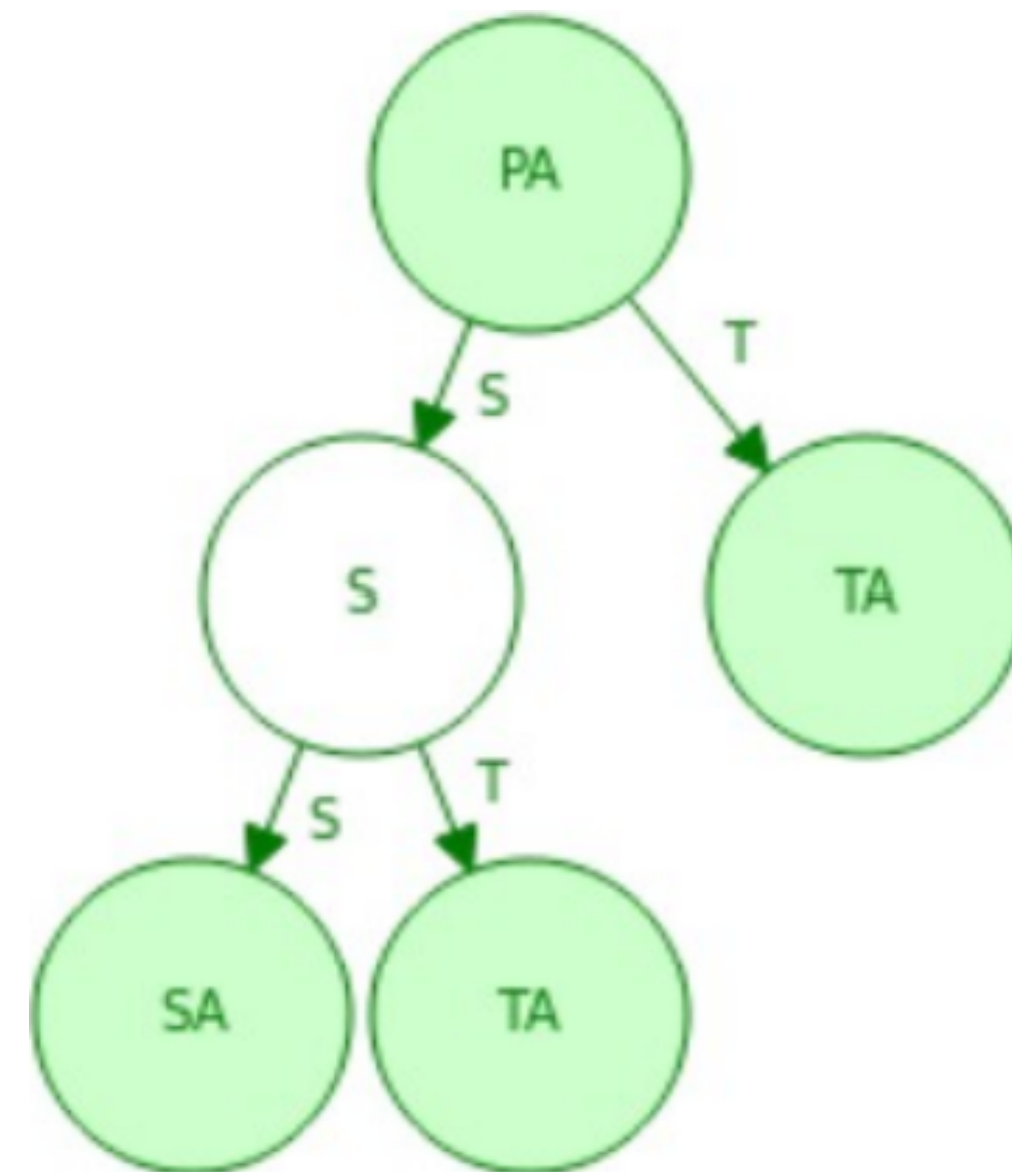
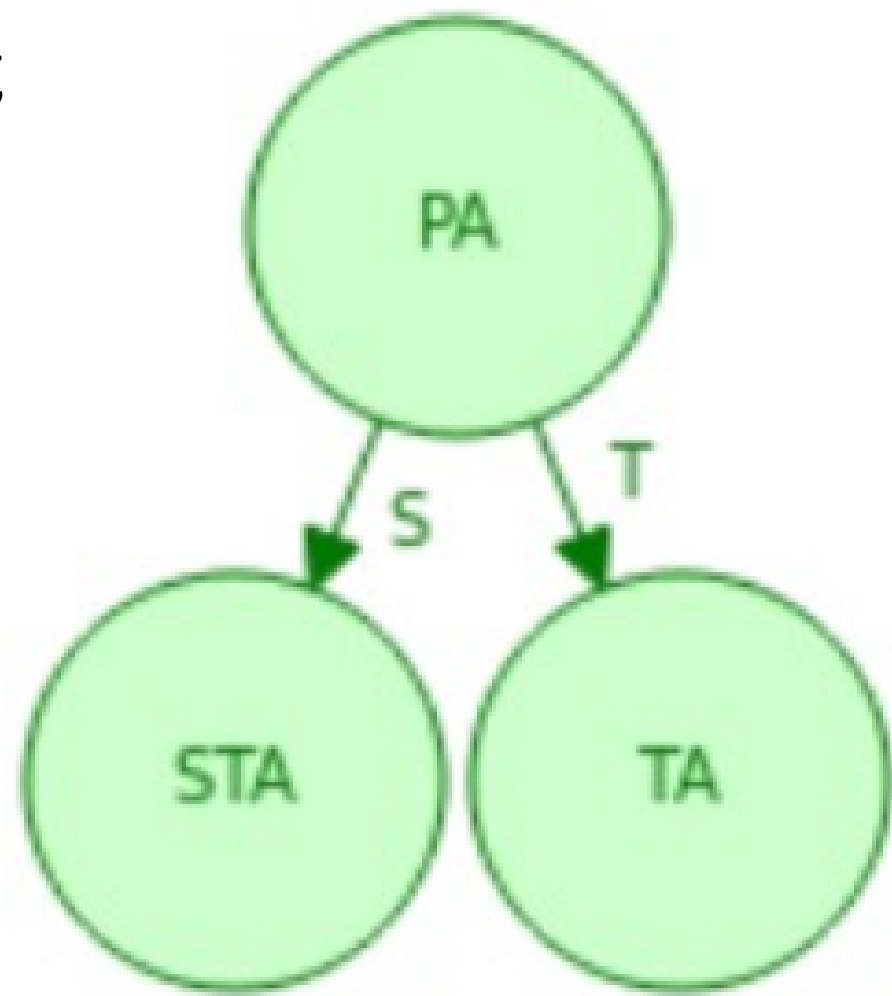
Construção da Árvore

- PASTA;



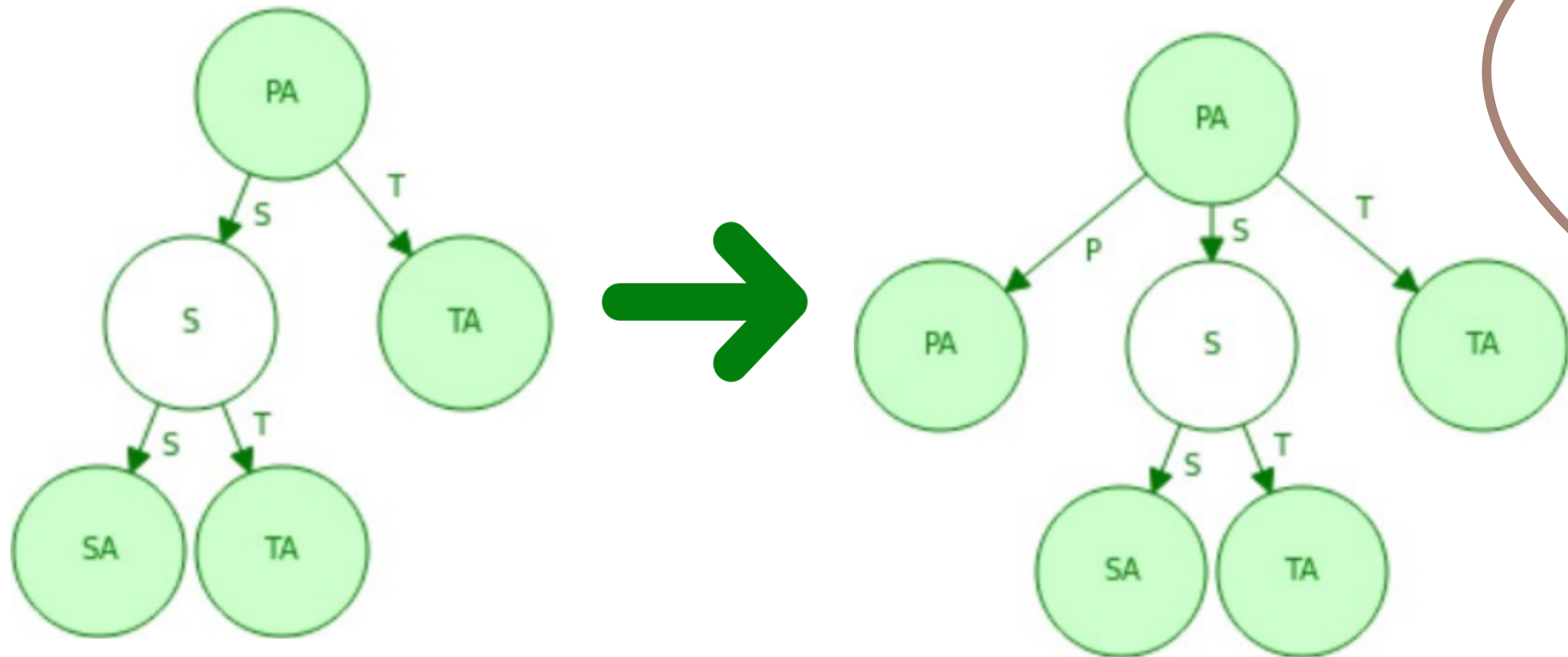
Construção da Árvore

- PASSA;



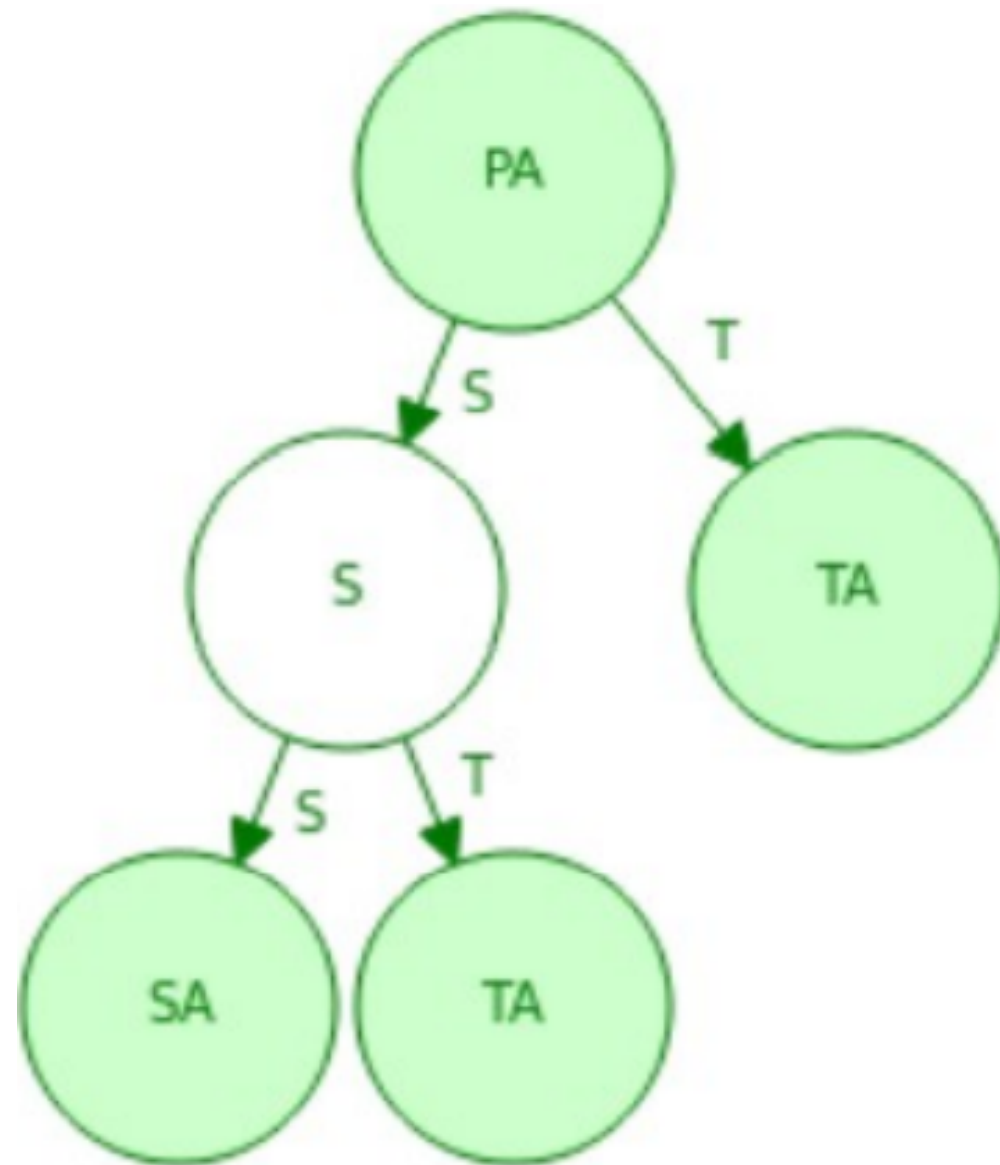
Construção da Árvore

- PAPA;

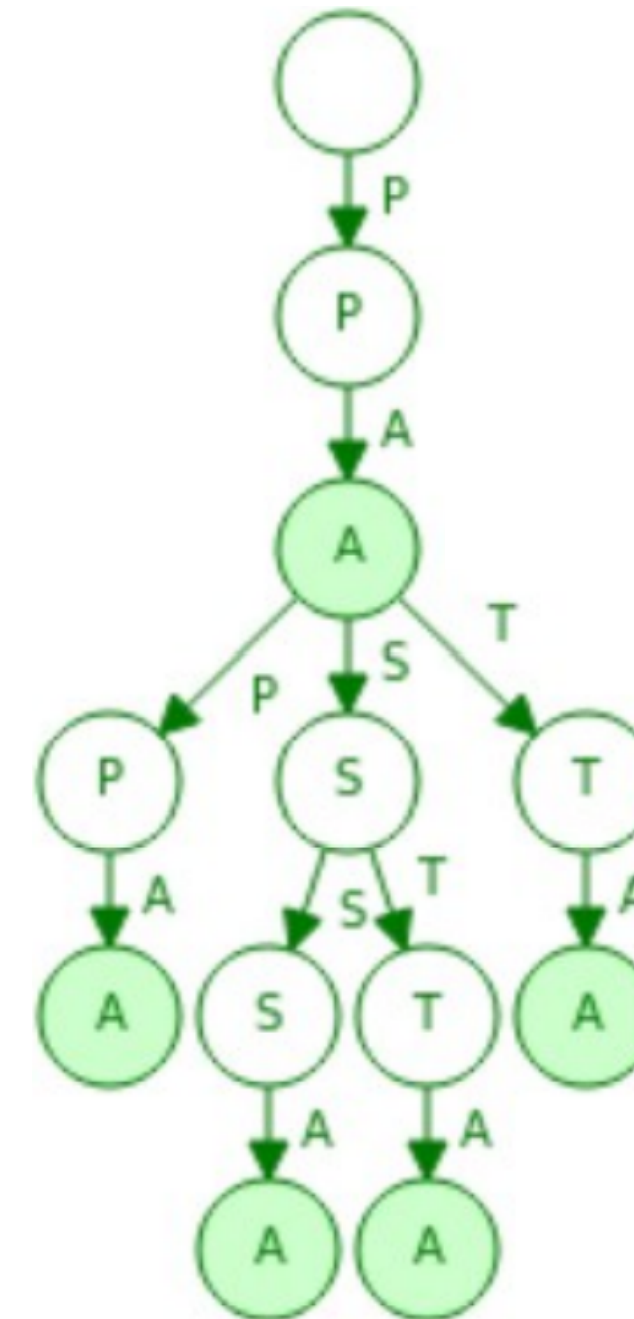


Comparando Árvores

- Patrícia:



- Trie:



Aplicação

- Dicionários;
- Corretores ortográficos;
- Programas para Processamento de Linguagem Natural (NLP – Natural Language Processing).

Portanto, como vimos, o algoritmo de busca PATRICIA ainda é muito utilizado atualmente, sendo usado em união à tabela hash.

Ao longo dos anos houve várias melhorias e a tendência é que cada vez mais seu uso seja aprimorado, a fim de atingir soluções rápidas e eficientes aos problemas futuros.

ÁRVORE PATRÍCIA

Obrigado!

Arthur, Sérgio e Vitória