Pesquisa digital

Algoritmos e Estruturas de Dados II

Pesquisa digital

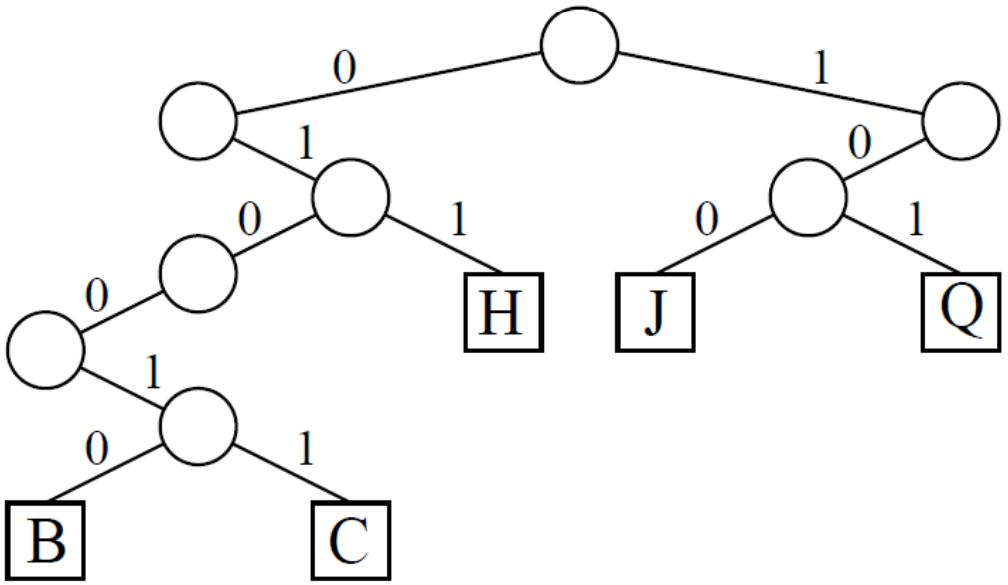
- A pesquisa digital usa a representação das chaves para estruturar os dados na memória
  - Por exemplo, a representação de um número em binário
  - A representação de um *string* com uma sequência de caracteres
- A pesquisa digital está para árvores binárias de pesquisa como radixsort está para os métodos de ordenação
  - Pesquisa não é baseada em comparação de chaves, mas sim em processamento feito sob a chave

Trie binária

- Cada nó no nível i representa o conjunto de todas as chaves que começam com a mesma sequência de i bits
- Cada nó tem duas ramificações, uma para as chaves cujo bit $(i+1)$ é zero e outra para chaves cujo bit $(i+1)$ é um

Trie binária – exemplo de pesquisa

- Busca em uma *trie* binária é parecida com pesquisa em árvore de busca
  - Mas não comparamos chaves
  - Percorremos a *trie* de acordo com os bits da chave



$$\begin{aligned} \mathrm{B} &= 010010 \\ \mathrm{C} &= 010011 \\ \mathrm{H} &= 011000 \\ \mathrm{J} &= 100001 \\ \mathrm{M} &= 101000 \end{aligned}$$

Trie – estruturas

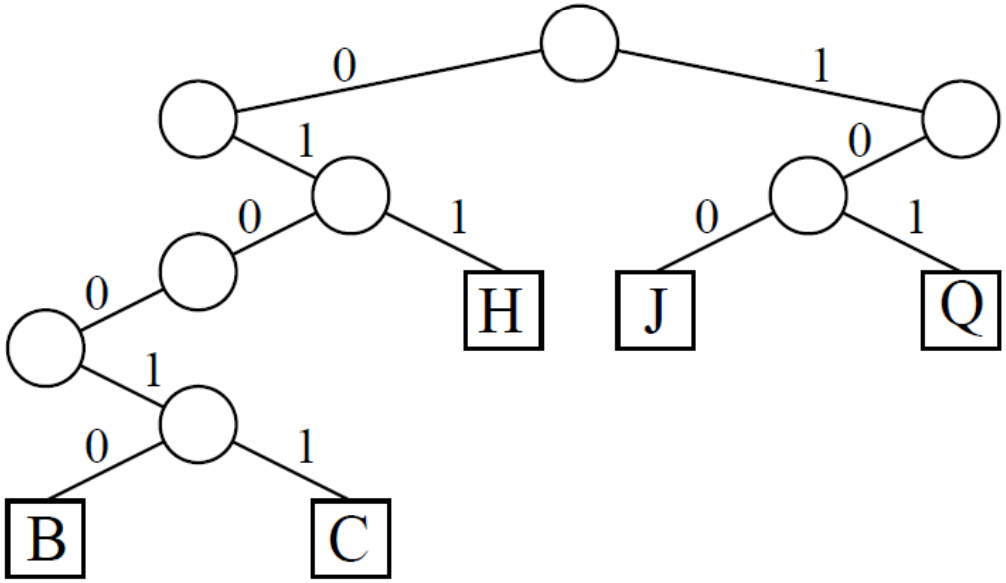
```
struct no {
    struct no *esq;
    struct no *dir;
    struct registro *reg;
};

struct registro {
    int chave;
    /* outros dados */
};
```

Trie binária – pesquisa

```
struct registro * pesquisaR(struct no *t, int chave,
                            int p) {
    if(t == NULL) return NULL;
    if(t->esq == NULL && t->dir == NULL) {
        int regchave = t->reg->chave;
        if(regchave == chave) { return t->reg; }
        else { return NULL; }
    }
    if(digit(chave, p) == 0) {
        return pesquisaR(t->esq, chave, p+1); }
    else { return pesquisaR(t->dir, chave, p+1); }
}
struct registro * pesquisa(struct no *trie, int chave)
{
    return pesquisaR(trie, chave, 0);
}
```

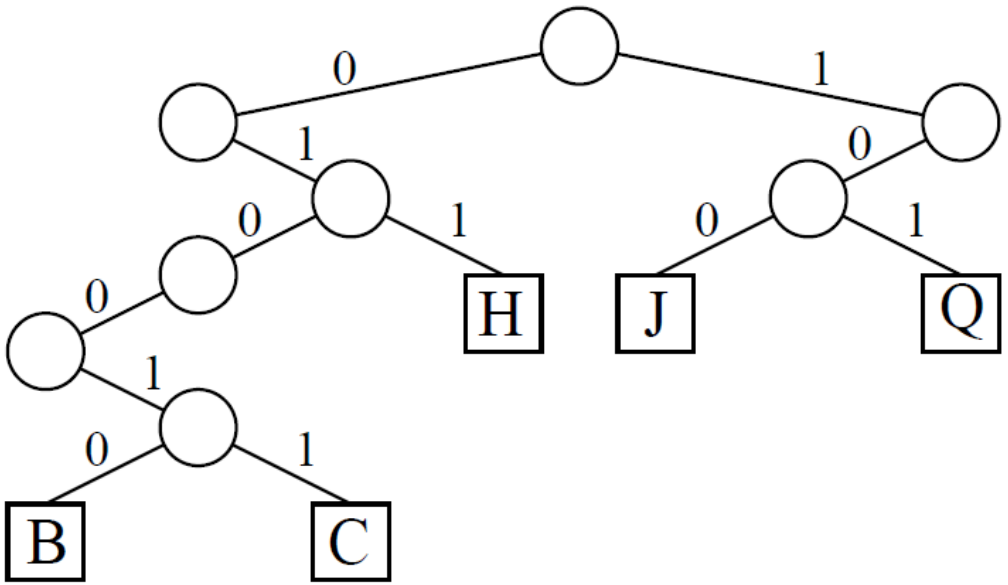
Trie binária – exemplo de inserção

- Fazemos uma pesquisa na árvore para descobrir onde a chave será inserida
  - Primeiro caso: se o nó externo onde a pesquisa terminar for fazio, basta cria um novo nó para conter a nova chave
  - Inserindo W = 110110



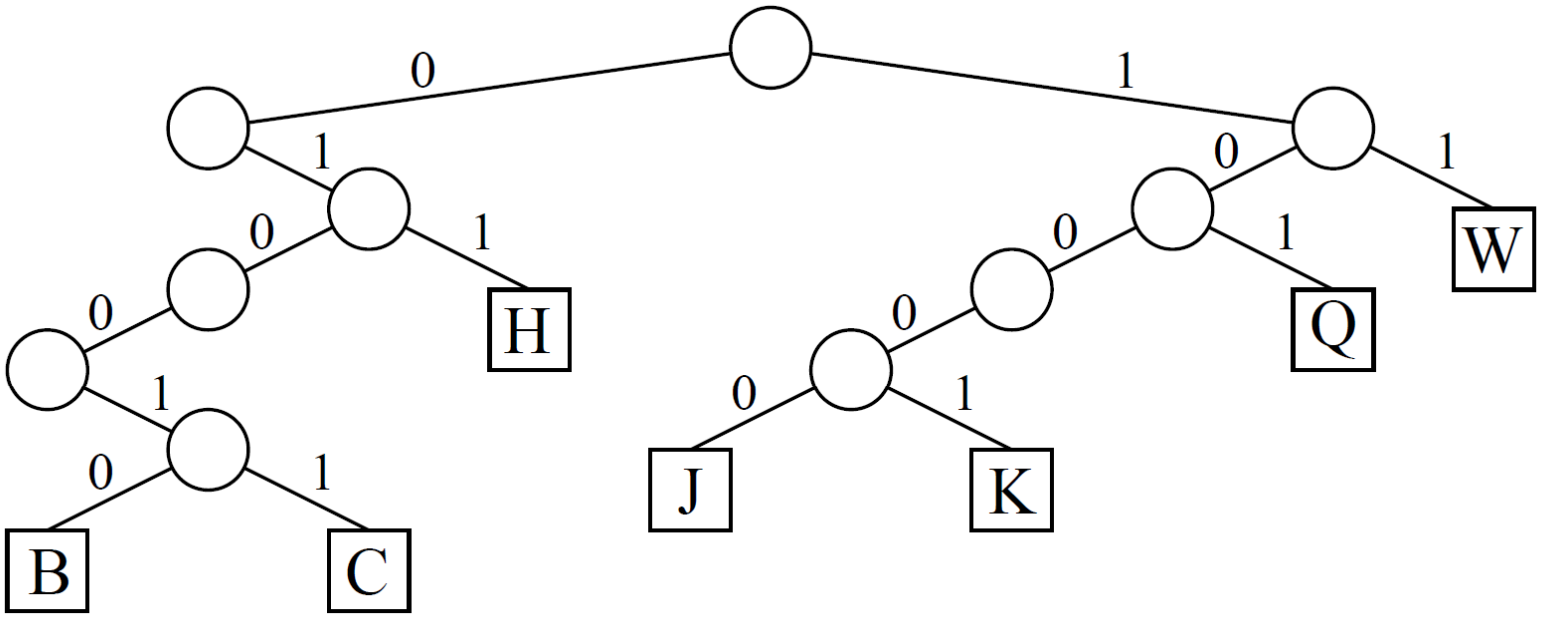
Trie binária – exemplo de inserção

- Fazemos uma pesquisa na árvore para descobrir onde a chave será inserida
  - Segundo caso: se o nó externo onde a pesquisa terminar tiver uma chave, criamos nós internos até encontrar o bit onde a nova chave difere da chave já existente
  - Inserindo K = 100010



$$\begin{aligned} \mathrm{B} &= 010010 \\ \mathrm{C} &= 010011 \\ \mathrm{H} &= 011000 \\ \mathrm{J} &= 100001 \\ \mathrm{M} &= 101000 \end{aligned}$$

Trie – exemplo de inserção



Trie – inserção

```
struct no * insereR(struct no *t, struct registro *reg,
                    int p) {
    int chave = reg->chave;
    if(t == NULL) return cria_trie(reg);
    if(t->esq == NULL && t->dir == NULL) {
        return separa(cria_trie(reg), t, p);
    }
    if(digit(chave, p) == 0)
        t->esq = insereR(t->esq, reg, p+1);
    else t->dir = insereR(t->dir, reg, p+1);
    return t;
}
void insere(struct no **trie, struct registro *reg) {
    *trie = insereR(*trie, reg, 0);
}
```

Trie – inserção

```
struct no * separa(struct no *novo, struct no *velho,
                   int p) {
    novo = cria_trie(NULL);
    int chave1 = no1->reg->chave;
    int chave2 = no2->reg->chave;
    if(digit(chave1, p) == 0 && digit(chave2, p) == 0){
        novo->esq = separa(no1, no2, p+1);
    } else if(/* 0 1 */) {
        novo->esq = no1; novo->dir = no2;
    } else if(/* 1 0 */) {
        novo->dir = no1; novo->esq = no2;
    } else if(/* 1 1 */) {
        novo->dir = separa(no1, no2, p+1);
    }
    return novo;
}
```

Vantagens

- O formato das *tries* **não** depende da ordem em que as chaves são inseridas
  - Depende apenas dos valores das chaves
  - “Balanceamento natural”
- Inserção e busca numa *trie* com N chaves aleatórias requer aproximadamente $\lg(N)$ comparações de bits no caso médio
  - Não depende do tamanho da chave
- O pior caso é limitado pelo número de bits das chaves

Desvantagens

- Caminhos de uma única direção acontecem quando chaves compartilham vários bits em comum
  - Por exemplo, as chaves B (00010) e C (00011) são idênticas exceto no último bit
  - Requer inspeção de todos os bits da chave independente do número de registros na *trie*
- Os registros são armazenados apenas nas folhas, o que desperdiça memória em nós intermediários
  - Uma *trie* com N registros tem aproximadamente $1.44N$ nós

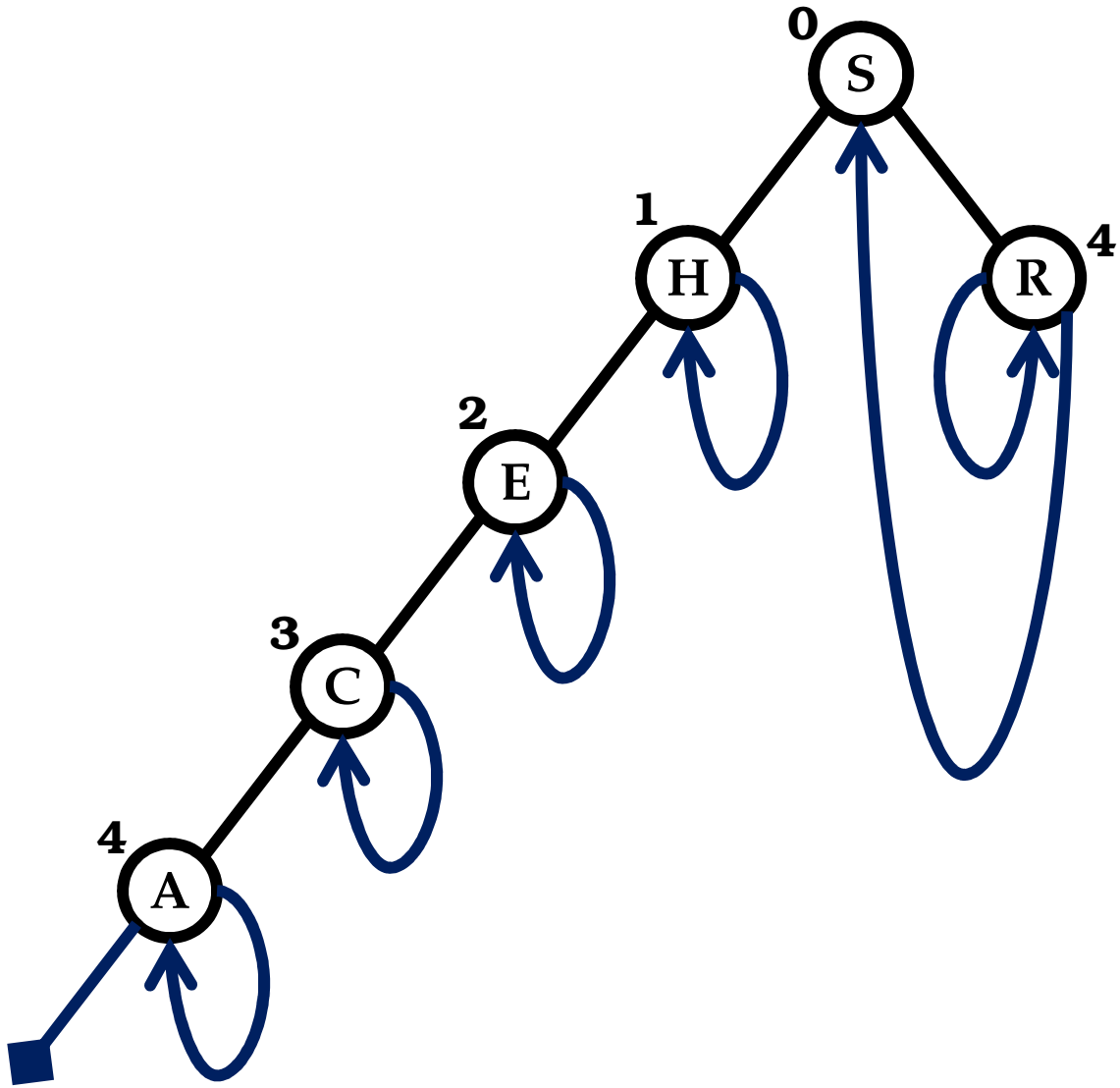
Patricia

- Practical Algorithm To Retrieve Information Coded in Alphanumeric
- Criada por Morrison 1968 para recuperação de informação em arquivos de texto
- Extendido por Knuth em 73, Sedgewick em 88, Gonnet e Baeza-Yates em 91

Patricia

- Patricia usa o conceito de pesquisa digital, mas estrutura os dados de forma a evitar as desvantagens citadas das *tries*
- Os problemas dos caminhos de única direção e desperdício de memória em nós internos é eliminado de forma elegante
- Cada nó da árvore contém uma chave e um índice indicando qual bit deve ser testado para decidir qual ramo seguir

Patricia – exemplo de pesquisa



Busca por
R = 10010
I = 01001

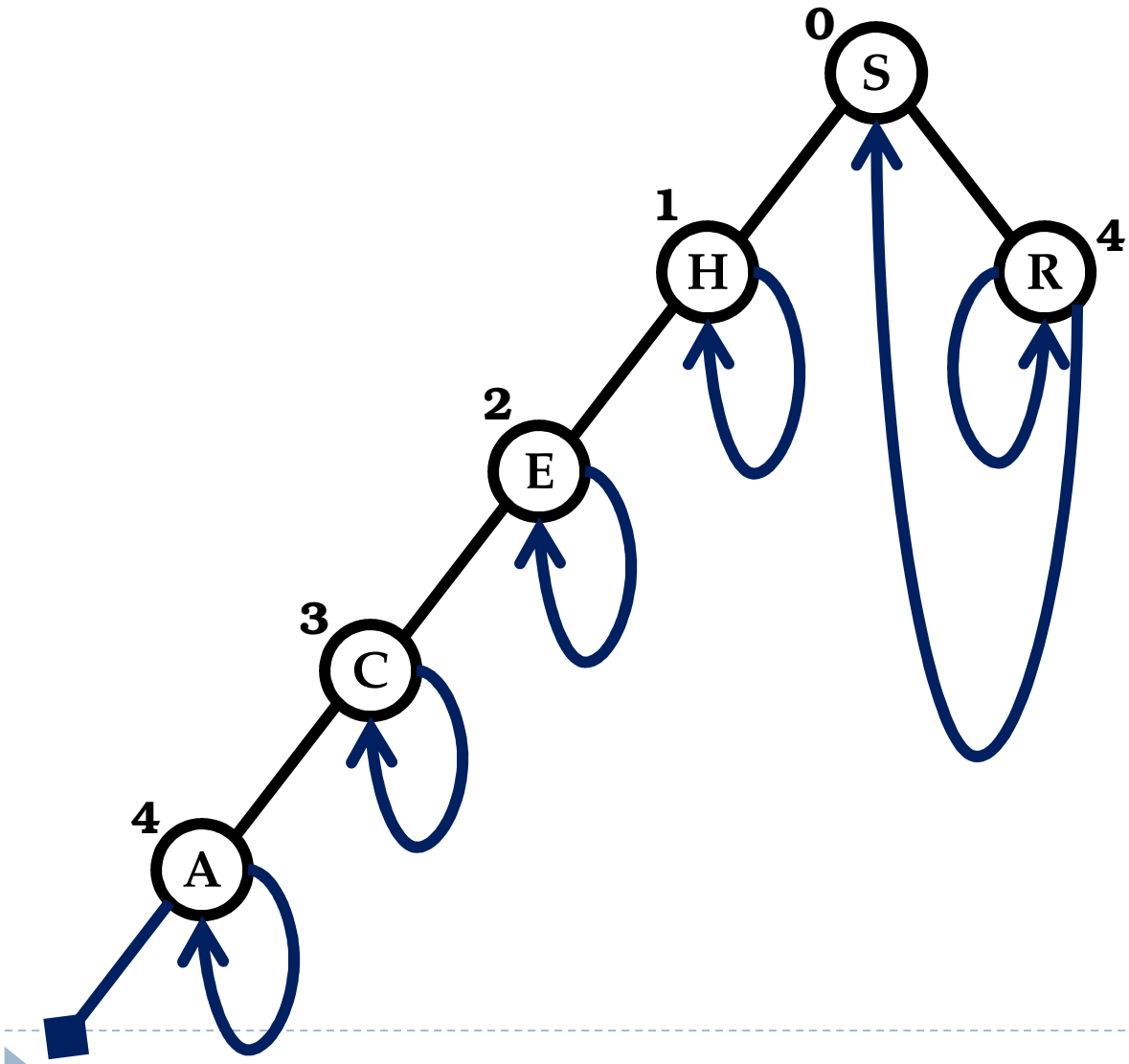
Patricia – estruturas

```
struct no {
    struct no *esq;
    struct no *dir;
    int bit;
    struct registro *reg;
};

struct registro {
    int chave;
    /* outros dados */
};
```

Patricia – pesquisa

```
struct registro * pesquisaR(struct no *t, int chave,
                            int bit) {
    if(t->bit <= bit) return t;
    if(digit(chave, t->bit) == 0) {
        return pesquisaR(t->esq, chave, t->bit);
    }
    else {
        return pesquisaR(t->dir, chave, t->bit);
    }
}
struct registro * pesquisa(struct no *pat, int chave) {
    struct registro *reg;
    reg = pesquisaR(pat->esq, chave, -1);
    if(reg->chave == chave) { return reg; }
    else { return NULL; }
}
```

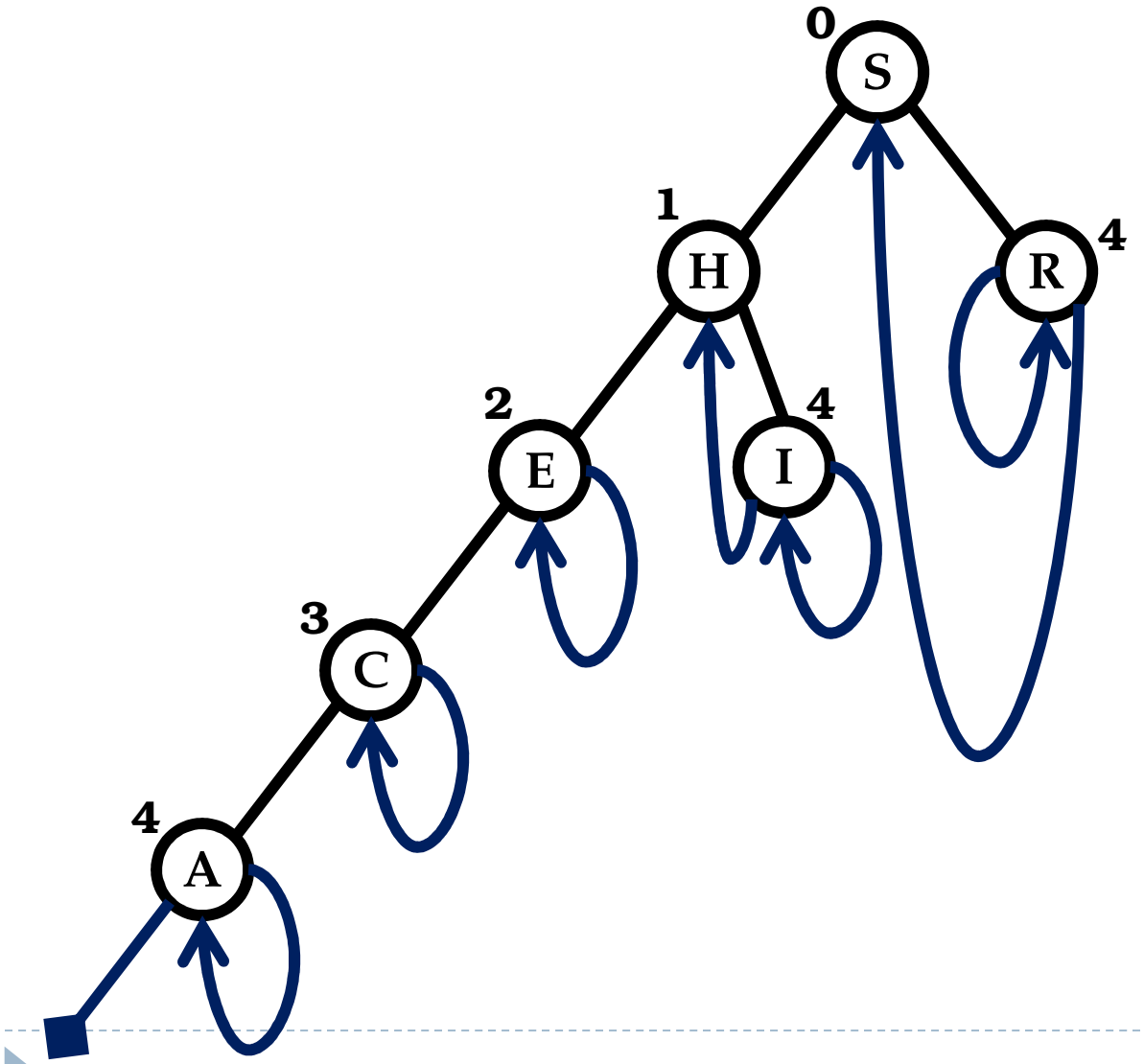
Patricia – exemplo de inserção



Inserindo
I = 01001

H = 01000

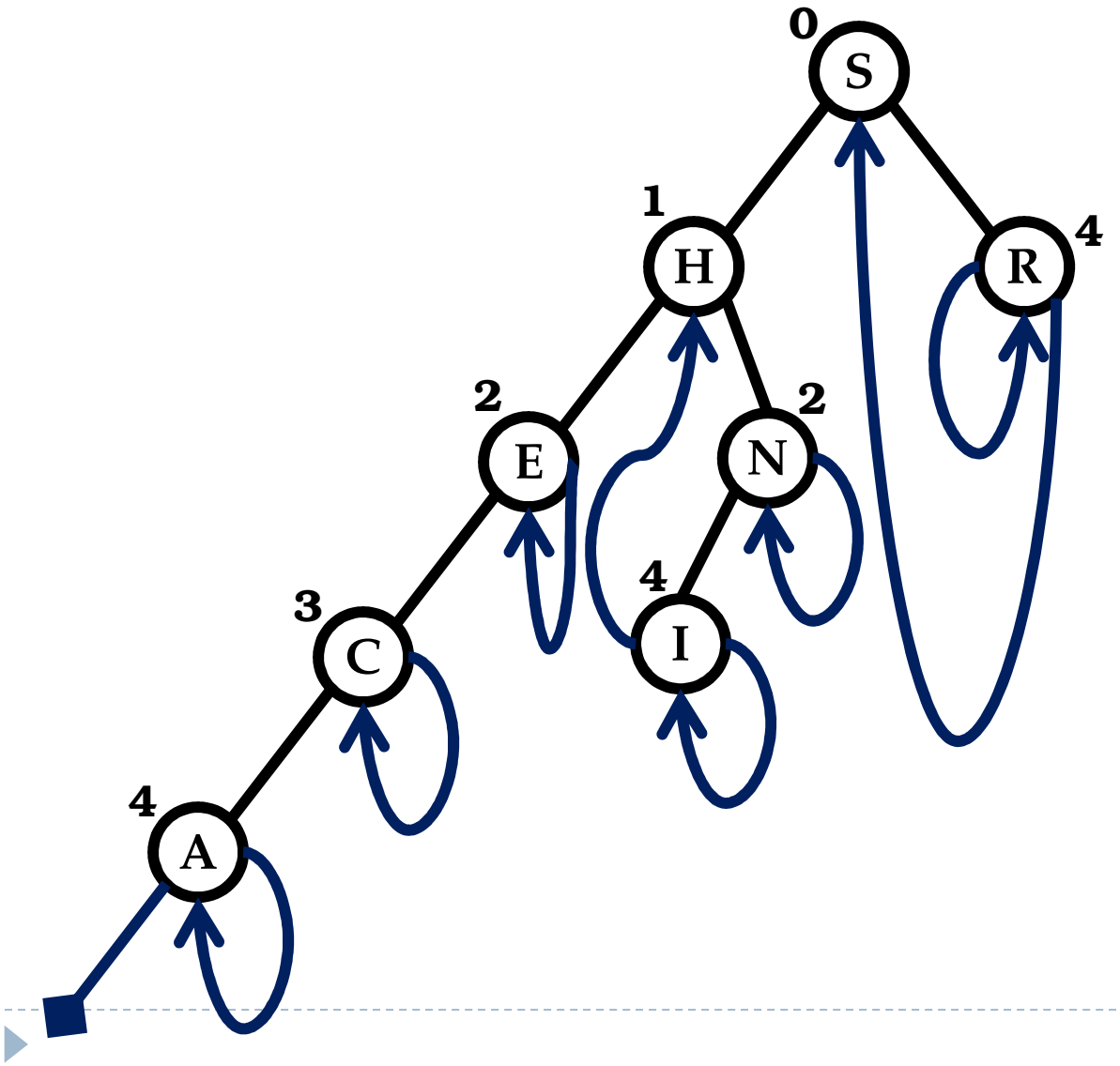
Patricia – exemplo de inserção (caso 1)



Inserindo
N = 01110

I = 01001

Patricia – exemplo de inserção (caso 2)



Casos da Inserção

- Caso 1: O novo nó substitui um nó externo
  - Acontece quando o bit que diferencia a nova chave da chave encontrada não foi utilizado na busca
- Caso 2: O novo nó substitui um nó interno
  - Acontece quando o bit que diferencia a nova chave da chave encontrada foi pulado durante a busca

Patricia – inserção

```
void insere(struct no *pat, struct registro *reg) {
    int chave = reg->chave;
    struct registro *ins = pesquisaR(pat, chave, -1);
    int ichave = ins->chave;
    if(chave == ichave) return;
    int i = 0;
    while(digit(chave, i) == digit(ichave, i)) i++;
    /* i é o bit diferenciador */
    pat->esq = insereR(pat->esq, reg, i, pat);
}
```

Patricia – inserção

```
struct no * insereR(struct no *t, struct registro *reg,
                    int bit, struct no *pai) {
    int chave = reg->chave;
    if((t->bit >= bit) || (t->bit <= pai->bit)) {
        struct no *x = cria_pat(reg, bit);
        x->esq = digit(chave, x->bit) ? t : x;
        x->dir = digit(chave, x->bit) ? x : t;
        return x;
    } else if(digit(chave, t->bit) == 0) {
        t->esq = insereR(t->esq, reg, bit, t);
    } else {
        t->dir = insereR(t->dir, reg, bit, t);
    }
    return t;
}
```

Considerações

- Inserção numa árvore patricia com N chaves aleatórias requer aproximadamente lg(N) comparações de bits no caso médio e 2lg(N) comparações no pior caso
- O número de comparações de bits é limitado pelo tamanho das chaves
- Não existe desperdício de memória nos nós internos
- Não existe caminhos de direção única

Tries M-árias

- Apresentamos apenas *tries* binárias mas podemos ter *tries* M-árias
- Cada filho corresponde a um possível valor do i-ésimo “bit”
  - Por exemplo, cada nó numa *trie* para armazenar *strings* pode ter 256 filhos (um pra cada valor possível para um caractere)

Exercícios

- Implemente as funções
  - struct no * cria_trie(struct registro *reg);
  - struct no * cria_pat(struct registro *reg, int bit);
- Mostre a *trie* resultante da inserção das chaves E X M P L O F A C
- Mostre a árvore patricia resultante da inserção das chaves E X M P L O F A C