Cours3
Parcours et recherche

Année 2022-2023 J-F. Kamp - IUT de Vannes

PLAN

- Parcours et recherche
  - Exemples
  - Recherche séquentielle
    - Première solution
    - Autres versions
  - Recherche dichotomique
    - Hypothèses
    - Illustration
    - Complexité

Parcours et recherche

Exemples

La recherche est un problème classique de tous les « jours » :

- rechercher un correspondant téléphonique
- rechercher un fichier dans un répertoire
- rechercher un mot-clé dans un document
- taper une requête sur Google

Exemples

Cette recherche se fait dans ce que l'on appelle une « collection » :

- collection de correspondants
- collection de fichiers
- collection de mots
- collection d'URLs

Exemples

La collection la + simple en ce qui nous concerne, c'est le tableau d'entiers de taille N.

29	50	-56	205	129	1000
0	1	2	3	4	5

Rechercher l'entier « 129 », c'est parcourir tout le tableau et s'arrêter lorsque l'on tombe sur la case contenant « 129 ».

Recherche séquentielle

Première solution

Hypothèses :

- Un tableau de taille N, rempli partiellement avec **n** entiers
- Le tableau contient AU PLUS un entier égale à la valeur à chercher

Boucle sur le tableau et comparaison des éléments un par un.

Conditions d'arrêt :

- Soit l'entier est trouvé (**trouve == true**)
- Soit on atteint le dernier élément du tableau (**i == n**)

Première solution

Conditions d'arrêt (deux) :

- Soit l'entier est trouvé (**trouve == true**)
- Soit on atteint le dernier élément du tableau (**i == n**)

Condition de continuation :

- Loi de Morgan

  non (A ou B) = non(A) et non(B)
- **(trouve == false) && (i != n) ou bien (i < n)**

En Java

En Java, méthode « rechercheSeq ».

```
/* *
   *  Cette méthode recherche une valeur dans un
   *  tableau d'entiers. Elle renvoie le numéro de la
   *  case contenant cet entier ou -1 si il n'existe pas.
   *  @param leTab le tableau des valeurs
   *  @param aRech valeur à rechercher
   *  @param n le nbre de valeurs
   *  @return le numéro de la case (ou -1)
   */
int rechercheSeq ( int [ ] leTab, int n, int aRech ) {
   // variables locales
   int i, ret ;
   boolean trouve ;
   // initialisations
   // boucle
   // terminaison
   return ret ;
}
```

En Java

En Java, méthode « rechercheSeqV1 ».

```
int rechercheSeqV1 ( int [ ] leTab, int n, int aRech ) {

    // variables locales
    int i, ret;
    boolean trouve ;

    // initialisations
    i = 0;
    trouve = false;
    ret = -1;

    // boucle
    while ( ( trouve == false ) && ( i < n ) ) {

        if ( aRech == leTab[i] ) {
            trouve = true;
            ret = i;
        }

        i++;
    }


    return ret ;
}
```

En Java

En Java, méthode « rechercheSeqV2 ».
Simplifions : pas besoin de « ret », …

```
int rechercheSeqV2 ( int [ ] leTab, int n, int aRech ) {

    // variables locales
    int i ;
    boolean trouve ;

    // initialisations
    i = 0;
    trouve = false;

    // boucle
    while ( !trouve && ( i < n ) ) {

        if ( aRech == leTab[i] ) {
            trouve = true;
        }

        else {
            i++;
        }
    }

    // terminaison
    if ( !trouve ) i = -1;

    return i ;
}
```

En Java

En Java, méthode « rechercheSeqV3 ». Simplifions : pas besoin de « ret », remplacer la condition « trouver » par break.

```
int rechercheSeqV3 ( int [ ] leTab, int n, int aRech ) {

    // variables locales
    int i ;

    // initialisations
    i = 0;

    // boucle
    while ( i < n ) {

        if ( aRech == leTab[i] ) {
            break;
        }

        i++;
    }

    // terminaison
    if ( i == n ) i = -1;

    return i ;
}
```

Analyse en complexité

Que vaut f (n) ?

- Dans le meilleur des cas ?

  La valeur se trouve en position zéro.

- Dans le pire des cas ?

  La valeur ne s'y trouve pas.

- Dans le cas moyen ?

  La valeur se trouve à la moitié du tableau (n/2).

Analyse en complexité dans le pire des cas

Hypothèses habituelles :

- Une déclaration : 1 opération élémentaire
- Une affectation : 1 opération élém.
- Une addition ou multiplication : 1 opération élém.
- Une incrémentation (i++) ou décrémentation (i--) : 2 opérations élém.
- Un test de comparaison (<=, <, >=, >, !=, ==) : 1 opération élém.
- Accès à 1 élément de tableau tab[i] : 0 opération élém.
- break : 0 opération élém.

Complexité de la version3

```
int rechercheSeqV3 ( int [ ] leTab, int n, int aRech ) {

    // variables locales
    int i ;                                  // 1 op
    // initialisations
    i = 0;                                   // 1 op
    while ( i < n ) {                        // 1 op

        if ( aRech == leTab[i] ) {           // 1 op
            break;                           // 0 op
        }

        i++;                                 // 2 op
    }
    if ( aRech != leTab[i] ) {               // 1 op
        i = -1;                              // 1 op
    }

    return i ;
}
```

- Avant la boucle : 1 op
- Init. de la boucle : 1 op
- Boucle effectuée (n – 1 + 1) = n fois
- Cond. continuation : 1 op
- Corps de la boucle : 1 op
- Incrémentation : 2 op
- Terminaison : 2 op (pire des cas)

Complexité de la version3 dans le pire des cas

Nbre opérations élémentaires = f (n)

= 1 + 1 + (n + 1) x 1 + (n x 3) + 2

= 5 + 4n

Soit n : la taille du problème.

L'algorithme est linéaire en **n**. On dira que l'algorithme est en **Θ(n)**.

Commentaires :

- Rien d'étonnant : 1 boucle => classe de complexité **n**.
- Essai d'amélioration de l'algorithme pour diminuer la constante **4** (devant n).

Recherche dichotomique

Hypothèse de départ

Le gain sur l'efficacité de la recherche repose sur une structure de départ particulière.

La collection dans laquelle on effectue la recherche est TRIEE.

Si elle est triée alors la recherche se fera beaucoup plus rapidement et sera < **Θ(n)**.

Exemple

-56	29	50	129	205	1000
0	1	2	3	4	5

Le tableau est trié par ordre croissant.

Supposons la recherche de « 129 ».

Au départ on « coupe » le tableau en deux parties « égales » : (0 + 5)/2 = 2 (partie entière).

-56	29	50	129	205	1000
0	1	2	3	4	5

Exemple

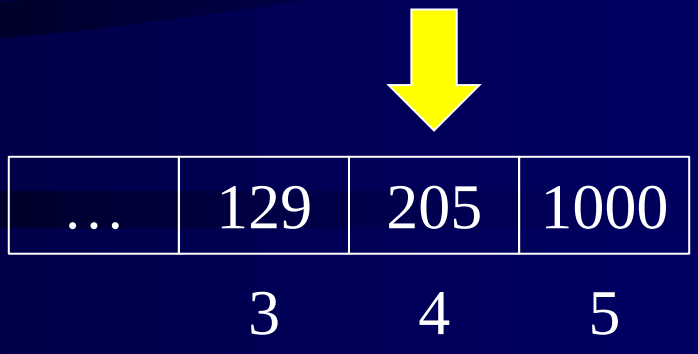
-56	29	50	129	205	1000
0	1	2	3	4	5

Comme tab[2] = 50 < 129, on est certain que « 129 » EST dans la partie du tableau à droite de l'indice 2 entre les indices 3 et 5.

La recherche se poursuit alors uniquement dans ce « nouveau » tableau.

…	129	205	1000
	3	4	5

Exemple



On recoupe le tableau en deux : (3 + 5)/2 = 4.

Comme tab[4] = 205 > 129, on est certain que « 129 » EST dans la partie du tableau à gauche de l'indice 4 entre les indices 3 et 3.

La recherche se poursuit alors uniquement dans ce « nouveau » tableau à 1 seule case => on a trouvé.

Exemple

-56	29	50	129	205	1000
0	1	2	3	4	5

Nombre de boucles pour trouver « 129 » ?

Soit « indM » l'indice milieu du sous-tableau :

- Etape 1 : indM = 2
- Etape 2 : indM = 4
- Etape 3 : indM = 3

Méthode séquentielle : 4 étapes

Est-ce vraiment intéressant ? Oui pour n 

L’algorithme

En Java, méthode « rechercheDicho».

```
int rechercheDicho ( int [ ] leTab, int n, int aRech ) {
    // variables locales
    // indD = indice de début du sous-tableau
    // indF = indice de fin du sous-tableau
    // indM = indice milieu entre indD et indF
    int indD, indF, indM, ret;

    // initialisations
    indD = 0;
    indF = n – 1;

    // boucle
    // terminaison
    return ret;
}
```

L’algorithme

```
int rechercheDicho ( int [ ] leTab, int n, int aRech ) {
    // variables locales
    …
    // initialisations
    indD = 0;
    indF = n – 1;
    // boucle
    while ( indD != indF ) {
        indM = ( indD + indF ) / 2;    // division entière !
        if ( aRech > leTab[indM] ) {
            indD = indM + 1;
        }
        else indF = indM;
    }
    // terminaison
    return ret;
}
```

L'algorithme

```
int rechercheDicho ( int [ ] leTab, int n, int aRech ) {
    // variables locales
    …
    // initialisations
    …
    // boucle
    while ( indD != indF ) {
        indM = ( indD + indF ) / 2;    // division entière !
        if ( aRech > leTab[indM] ) {
            indD = indM + 1;
        }
        else indF = indM;
    }
    // terminaison : indD == indF forcément
    // MAIS leTab [indD] par forcément = aRech !!
    if ( aRech == leTab [indD] ) ret = indD;
    else ret = -1;
    return ret;
}
```

Analyse en complexité

Que vaut f (n) ?

- Dans le meilleur des cas ?
- Dans le pire des cas ?
- Dans le cas moyen ?

Analyse en complexité

```
int rechercheDicho ( int [ ] leTab, int n, int aRech ) {
    int indD, indF, indM, ret;                          // 4 op

    indD = 0;                                           // 1 op
    indF = n – 1;                                       // 2 op

    while ( indD != indF ) {                            // 1 op
        indM = ( indD + indF ) / 2;                     // 3 op
        if ( aRech > leTab[indM] ) {                    // 1 op
            indD = indM + 1;                            // 2 op
        }
        else indF = indM;                               // 1 op
    }
    if ( aRech == leTab [indD] ) ret = indD;            // 2 op
    else ret = -1;                                      // 1 op

    return ret;
}
```

Analyse en complexité

Combien de fois la boucle while (indD != indF) est exécutée ?

Réponse : tant que la taille tableau > 1

Supposons aRech se trouve en première case (= case 1 par convention = indD).

Au départ taille tableau : indF – indD + 1 = n

Itération **1** : indM = (n + 1) / 2 ≈ n / 2^{1}

taille : indM – indD + 1= n / 2^{1}

Itération **2** : indM = n / 4 = n / 2^{2}

taille : indM – indD + 1= n / 2^{2}

Itération **3** : indM = n / 8 = n / 2^{3}

taille : indM – indD + 1= n / 2^{3}

… …

Itération **k** : indM = n / 2^{k}

taille : indM – indD + 1 = 1

Analyse en complexité

Combien de fois la boucle while (indD != indF) est exécutée ?

k est le nombre de fois que la boucle est exécutée.

Itération **k** : indM = n / $2^{\mathbf{k}}$ = 1

Que vaut **k** = ?

Reste à résoudre : 1 = n / $2^{\mathbf{k}}$

$2^{\mathbf{k}}$ = n

=> $k = \log_2 n$

Complexité de la recherche dichotomique

Nbre opérations élémentaires dans le pire des cas = f (n)

nbTours = nbT = **$k = \log_2 n$**

avant la boucle : 4 + 1 + 2 = 7

corps de boucle (pire cas) : (3 + 1 + 2) X nbT

cond. de continuation : 1 X (nbT + 1)

après la boucle : 2

f(n) = 7 + 6 nbT + 1 nbT + 1 + 2

f(n) = 10 + 7 x **k**

f(n) = 10 + 7 **$\log_2(n)$**

Complexité de la recherche dichotomique

L'algorithme est logarithmique en **n**. On dira que l'algorithme est en **$\Theta(\log_2 n)$**.

Commentaires :

Intéressant surtout pour n ↗

Soit $n = 2^{20} = 1.048.576$

Recherche séquentielle :

$f(n) = 5 + 4n \approx$ **4.000.000**

Recherche dichotomique :

$f(n) \approx 7 \log_2(n) =$ **140** (!!)