

Stage ETE 2015

Description du système

F-PHT

Un système d'index de filtres de Bloom pour la recherche
d'information par mots clés

Réalisé par **DOAN** Cao Sang
Encadrant : M. **MAKPANGOU** Mesaac, Regal

30 Juin 2015

Table des matières

1	Vue globale	2
1.1	Prefix Hash Tree (PHT)	2
1.2	Fragment	4
1.3	F-PHT	5

Chapitre 1

Vue globale

1.1 Prefix Hash Tree (PHT)

Un arbre préfixe est un arbre numérique ordonné qui est utilisé pour stocker une table associative où les clés sont généralement des chaînes de caractères. Contrairement à un arbre binaire de recherche, aucun nœud dans le trie ne stocke la chaîne à laquelle il est associé. C'est la position du nœud dans l'arbre qui détermine la chaîne correspondante¹.

Pour tout nœud, ses descendants ont en commun le même préfixe. La racine est associée à la chaîne vide. Des valeurs ne sont pas attribuées à chaque nœud, mais uniquement aux feuilles et à certains nœuds internes se trouvant à une position qui désigne l'intégralité d'une chaîne correspondant à une clé.

1. Wikipédia

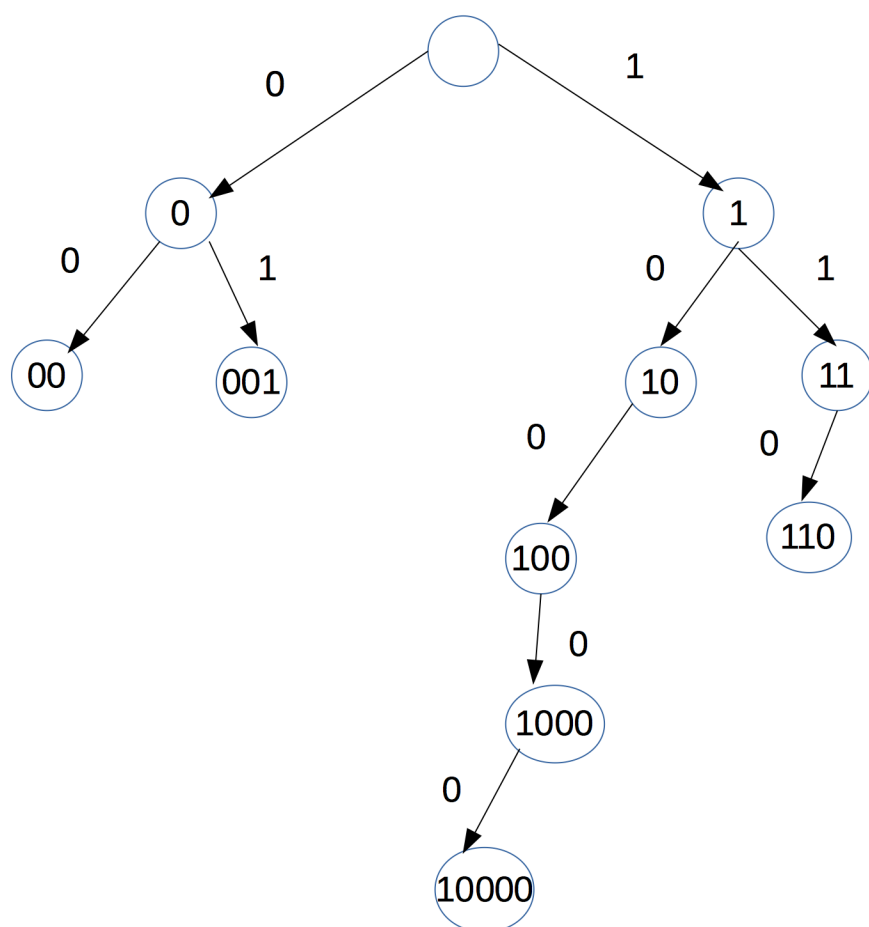


FIGURE 1.1 – Un arbre préfixe

1.2 Fragment

On considère un filtre de Bloom de taille m . Le système découpe ce filtre en f fragments de taille identique. Par convention, les fragments sont numérotés de 0 à $f-1$, en commençant par le fragment le plus à gauche. L'identifiant de chaque fragment est défini de façon unique.

Par exemple, $m = 16$, $f = 4$.

15												0			
1	0	0	0	1	1	0	1	0	0	0	0	1	0	1	0

TABLE 1.1 – Exemple le filtre de Bloom

1	0	0	0
---	---	---	---

TABLE 1.2 – Exemple le fragment $f = 0$

1	0	1	0
---	---	---	---

TABLE 1.3 – Exemple le fragment $f = 3$

1.3 F-PHT

F-PHT est une sorte d'arbre préfixe, il utilise les filtres de Bloom de taille m comme clés de stockage. Cet arbre utilise l'identifiant d'un fragment de filtre de Bloom à la place de préfixe. Chaque nœud de l'arbre stocke un ensemble de couple $\langle \text{prefix}, \text{identifiant} \rangle$ avec **prefix** est la valeur d'un fragment de rang **i** et **identifiant** est soit l'identifiant d'un nœud soit null. Si **identifiant** est égal à null, **prefix** est un filtre de Bloom. Sinon, ce couple désigne où sont stockés les filtres de Bloom ayant **prefix** comme valeur du fragment de rang **i**, où **i** correspond au niveau de ce nœud dans l'arbre.

Chaque nœud contient une liste **RouteEntry**, qui contient les couples stockés par ce nœud. Chaque nœud stocke au plus γ couples, avec γ inférieur ou égal à $2^{\frac{m}{f}}$