# Kap. 5 Tries, digitale Suchbäume, Sorted frequency trees

**Tries**: Trie ist Spezialfall von Positionsbaum für Text = Menge von Wörtern, die mit *blank* abgeschlossen sind, d.h.

- keineWiederholungen
- kein Wort ist Präfix eines anderen.
- Info am Blatt: Rest des Wortes, sonstige Information, z.B. Stamm, Etymologie, Synonyma, ..., Bedeutung

**Notation**:

f(k) : Häufigkeit des Wortes in Knoten k
h(k) : Höhe des Knotens k von Wurzel gezählt
f(k) • h(k) : Anzahl Suchschritte

1

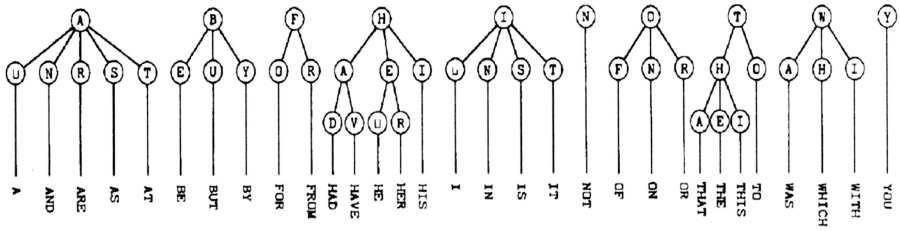



**Fig. 31.** The trie of Table 1, converted into a "forest."

2

## Übergang zu digitalem Suchbaum:

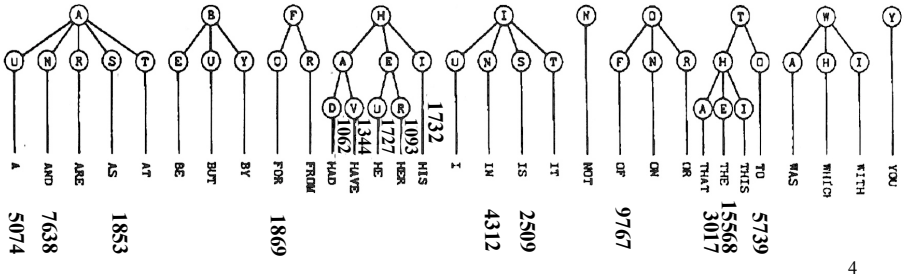
**Idee:** wie Trie, aber häufigstes Wort jedes Unterbaums in Wurzel!

Für 10 häufigste Wörter:

$$\sum f(k)h(k) = 190623 \quad \text{im Trie}$$

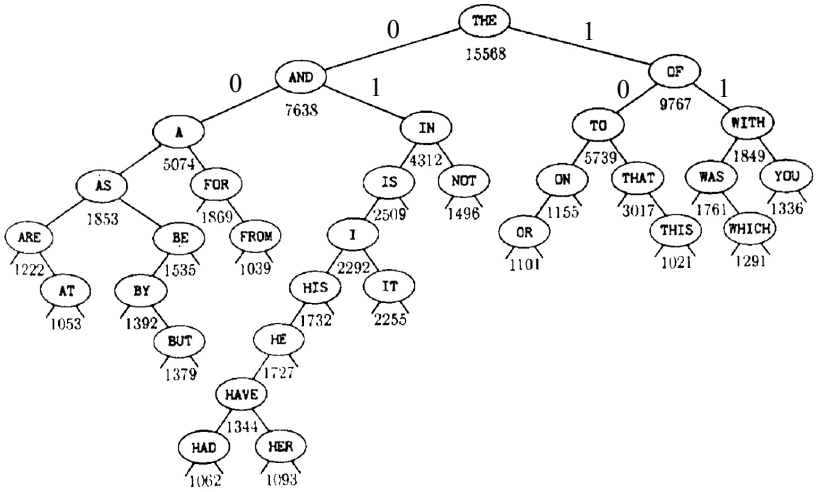
$$\sum f(k)h(k) = 108560 \quad \text{im dig. Suchbaum}$$

3




4

## Digitaler Suchbaum für Binärform von Wörtern, häufigstes Wort in Wurzel




A digital search tree for the 31 most common English words, insted in decreasing order of frequency.

5

**Sorted frequency tree**

| | |
|---|---|
| THE | 15568 |
| OF | 9767 |
| AND | 7638 |
| TO | 5739 |
| A | 5074 |
| IN | 4312 |
| THAT | 3017 |
| IS | 2509 |
| FOR | 1869 |
| AS | 1853 |

- einfügen nach absteigender Häufigkeit
- Knoten-Inhalt bestimmt Such- und Einfügpfad

6

**Vergleich**:

**1. Trie**: eindeutig, Info nur in Blättern, sortiert

**2. Digitaler Suchbaum:**
- eindeutig modulo gleichhäufiger Wörter
- Info in Zwischenknoten
- Wurzel enthält häufigstes Wort im Baum
- Baum nicht sortiert, aber
  $\forall x \in LB \;\; \forall y \in RB \quad : x \leq y$
- Verarbeitung, Ausgabe in Sortierreihenfolge?

**3. Sorted frequency tree:**
- Struktur abhängig von Einfügereihenfolge bei gleichhäufigen Wörtern
- sortiert

7