The talk will consider aspects of the following setup: Assume for each (key) $x$ from a set $U$ (the universe) a vector $a_x = (a_{x,0}, \ldots, a_{x,m-1})$ has been chosen. Given a list $Z = (z_i)_{i \in [m]}$ of vectors in $\{0, 1\}^r$ we obtain a mapping

$$\varphi_Z : U \to \{0, 1\}^r, x \mapsto \langle a_x, Z \rangle := \bigoplus_{i \in [m]} a_{x,i} \cdot z_i,$$

where $\bigoplus$ is bitwise XOR. The simplest way for creating a data structure for calculating $\varphi_Z$ is to store $Z$ in an array $Z[0..m-1]$ and answer a query for $x$ by returning $\langle a_x, Z \rangle$. The length $m$ of the array should be $(1 + \varepsilon)n$ for some small $\varepsilon$, and calculating this inner product should be fast.

In the focus of the talk is the case where for all or for most of the sets $S \subseteq U$ of a certain size $n$ the vectors $a_x, x \in S$, are linearly independent. Choosing $Z$ at random will lead to hash families of various degrees of independence. We will be mostly interested in the case where $a_x, x \in U$ are chosen independently at random from $\{0, 1\}^m$, according to some distribution $D$. We wish to construct (static) retrieval data structures, which means that $S \subseteq U$ and some mapping $f : S \to \{0, 1\}^r$ are given, and the task is to find $Z$ such that the restriction of $\varphi_Z$ to $S$ is $f$. For creating such a data structure it is necessary to solve the linear system

$$(a_x)_{x \in S} : Z = (f(x))_{x \in S}$$

for $Z$. Two problems are central: Under what conditions on $m$ and $D$ can we expect the vectors $a_x, x \in S$ to be linearly independent, and how can we arrange things so that in this case the system can be solved fast, in particular in time close to linear (in $n$, assuming a reasonable machine model)? Solutions to these problems, some classical and others that have emerged only in recent years, will be discussed.

2012 ACM Subject Classification Theory of computation $\rightarrow$ Sorting and searching; Theory of computation $\rightarrow$ Randomness, geometry and discrete structures

Keywords and phrases Hashing, Retrieval, Linear equations, Randomness

Digital Object Identifier 10.4230/LIPIcs.ESA.2023.1

Category Invited Talk