Appendix

A Notations

All notations in the main text and their descriptions are summarized in Table 5.

Notation	Description
$x, \mathcal{X}$	Query, Query set
$y, \mathcal{Y}$	Item, Item Set
$\mathcal{X}_y^+$	Queries related to item y
$\mathcal{Y}_x^+$	Items related to query x
$\mathcal{D}$	A Dataset consisting of $\mathcal{X}$ and $\mathcal{Y}$
$\mathcal{Q}_x$	The distribution of items related to x in the index
p	Item-query relevance probability
m	The number of queries
N	The number of items
δ	A tolerance level
$\Phi(\cdot)$	The CDF of standard normal distribution
P_Q^i	The distribution at the i-th token of the query
P_I^i	The distribution at the i-th token of the item
C^i	The token assigned to the item at the i-th Layer
$H(a, b)$	Cross-entropy of a and b
$H(b)$	Entropy of b
U^k	A uniform distribution over k classes
k	The width of the index (The number of tokens)
L	The depth of the index (The number of Layers)

Table 5: Notations

B Dataset Statistics

Dataset	#Query	#Item	# Interaction
KuaiSAR	191330	112388	1093920
Beauty	41895	15817	162713
Toys and Games	13271	25357	90557

Table 6: Dataset Statistics

We summarize the statistics of the three datasets in Table 6. Setting the tolerance level δ to 0.95 and substituting the dataset density for p, all three datasets satisfy the conditions of Theorem 1. This indicates that our proposed greedy algorithm is applicable to most datasets and demonstrates the rationality of URI's approach in simulating the greedy algorithm through machine learning.

C Proof of Theorem 1

Before deriving Theorem 1, we can transform the problem of recovering the allocation scheme into a matrix equation-solving problem. Let W be the interaction matrix between queries and items, with a size of $m \times N$. The value at a given position is 1 if the query is related to the item, and 0 otherwise. In a recommender system, W represents the user purchase matrix. Let H be the item allocation matrix, with a size of $N \times k$, where each row contains exactly one 1, and the sum of each column is N/k. Define $M = W \times H$, where M has a size of $m \times k$, representing the frequency distribution of each query over the k buckets. Thus, the problem is transformed into solving for H given M and W. The conclusions of the theorem and the greedy algorithm are also correspondingly transformed, as described below.

C.1 THEOREM

Theorem 2. *(Correct Reconstruction Probability)*

Given the following conditions:

- *Tolerance Level: A tolerance level $\delta \in (0, 1)$.*
- *Parameters: Positive integers N and k, and a probability $p \in (0, 1)$.*
- *Random Matrix W: An $m \times N$ binary matrix where each element $W_{r,l}$ is independently and identically distributed (i.i.d.), satisfying:*

$$P(W_{r,l} = 1) = p, \quad P(W_{r,l} = 0) = 1 - p$$

- *Structured Matrix H: An $N \times k$ binary matrix satisfying:*
  - *Row Constraint: Each row contains exactly one 1, and all other elements are 0.*

$$\sum_{j=1}^{k} H_{i,j} = 1, \quad \forall i = 1, 2, \ldots, N$$

  - *Column Constraint: Each column contains exactly $N_j = \frac{N}{k}$ ones (assuming N is divisible by k).*

$$\sum_{i=1}^{N} H_{i,j} = N_j, \quad \forall j = 1, 2, \ldots, k$$

- *Product Matrix M: Defined as $M = W \cdot H$.*

If the number of rows m in matrix W satisfies:

$$m \geq \frac{\left[\Phi^{-1}(1-\delta)\right]^2 \left(1 + \frac{2N}{k} p(1+p)\right)}{p(1-p)}$$

Then, using the following greedy algorithm, the probability that each row of matrix H is correctly reconstructed is:

$$P_{correct} \geq 1 - \delta$$

C.2 GREEDY ALGORITHM DESCRIPTION

For each row index $i = 1, 2, \ldots, N$:

1. Identify Set S_i: Find all rows in matrix W where the i-th column is 1.
$$S_i = \{r \mid W_{r,i} = 1\}$$
2. Compute Vector s_i: For each column $j = 1, 2, \ldots, k$, calculate the sum of corresponding entries in matrix M:
$$s_i^{(j)} = \sum_{r \in S_i} M_{r,j}$$
3. Select Maximum: Determine the column j^* with the highest sum:
$$j^* = \arg\max_j s_i^{(j)}$$
4. Update Matrix H: Assign the 1 in the i-th row of H to column j^*:
$$H_{i,j^*} = 1$$

C.3 Proof

Step 1: Definition of Difference Variable $\Delta s_i^{(j)}$

For a given row index i and any incorrect column $j \neq j^*$ (where j^* is the correct column), define the difference:

$$\Delta s_i^{(j)} = s_i^{(j^*)} - s_i^{(j)}$$

Using the structure of the problem:

$$\Delta s_i^{(j)} = c_{i,i} + \sum_{l \neq i} (H_{l,j^*} - H_{l,j}) c_{i,l}$$

where: $c_{i,i}$ is the number of rows in matrix W where column i has a value of 1. $c_{i,l}$ is the number of rows in matrix W where both columns i and l have a value of 1.

Step 2: Calculate Expectation and Variance of $c_{i,i}$ and $c_{i,l}$

(a) For $c_{i,i}$:

Since $c_{i,i}$ counts the number of times the i-th column in W contains a 1, it can be written as:

$$c_{i,i} = \sum_{r=1}^{m} W_{r,i}$$

where $W_{r,i}$ is an independent Bernoulli random variable with probability p. Therefore:

- Expectation:

$$E[c_{i,i}] = m \cdot p$$

- Variance:

$$\text{Var}[c_{i,i}] = m \cdot p(1-p)$$

(b) For $c_{i,l}$ (when $l \neq i$):

Similarly, $c_{i,l}$ counts the number of times both columns i and l in matrix W contain a 1. Since $W_{r,i}$ and $W_{r,l}$ are independent Bernoulli variables:

- Expectation:

$$E[c_{i,l}] = m \cdot p^2$$

- Variance:

$$\text{Var}[c_{i,l}] = m \cdot p^2(1-p^2)$$

Step 3: Expectation and Variance of $\Delta s_i^{(j)}$

We now calculate the expectation and variance of $\Delta s_i^{(j)}$.

(a) Expectation of $\Delta s_i^{(j)}$:

Using the linearity of expectation and the fact that the expectation of $H_{l,j^*} - H_{l,j}$ is zero (as both are binary variables):

$$E[\Delta s_i^{(j)}] = E[c_{i,i}] = m \cdot p(1-p)$$

(b) Variance of $\Delta s_i^{(j)}$:

Since $c_{i,i}$ and $c_{i,l}$ are independent, the variance of $\Delta s_i^{(j)}$ is the sum of the variances of the terms:

$$\text{Var}[\Delta s_i^{(j)}] = \text{Var}[c_{i,i}] + \sum_{l \neq i} (H_{l,j^*} - H_{l,j})^2 \cdot \text{Var}[c_{i,l}]$$

Since $(H_{l,j^*} - H_{l,j})^2$ takes values of 0 or 1 (with 1 occurring when one column contains a 1 and the other does not), let n be the number of times $(H_{l,j^*} - H_{l,j})^2 = 1$. We have:

$$n \leq \left(\frac{N}{k} - 1 \right) + \frac{N}{k} \leq \frac{2N}{k}$$

Thus, the variance becomes:

$$\mathrm{Var}[\Delta s_i^{(j)}] = \mathrm{Var}[c_{i,i}] + n \cdot \mathrm{Var}[c_{i,l}]$$

Substitute the values of $\mathrm{Var}[c_{i,i}]$ and $\mathrm{Var}[c_{i,l}]$:

$$\mathrm{Var}[\Delta s_i^{(j)}] = m \cdot p(1-p) + n \cdot m \cdot p^2(1-p^2)$$

Step 4: Apply the Central Limit Theorem

By the Central Limit Theorem, since m is a fairly large number, we approximate $\Delta s_i^{(j)}$ as a normally distributed variable:

$$\Delta s_i^{(j)} \sim \mathcal{N}\left(E[\Delta s_i^{(j)}], \mathrm{Var}[\Delta s_i^{(j)}]\right)$$

Thus, the probability of correct assignment is:

$$P\left(\Delta s_i^{(j)} > 0\right) = \Phi\left(\frac{E[\Delta s_i^{(j)}]}{\sqrt{\mathrm{Var}[\Delta s_i^{(j)}]}}\right)$$

where Φ is the CDF of the standard normal distribution.

Step 5: Substitute the Expectation and Variance

Substitute the expressions for $E[\Delta s_i^{(j)}]$ and $\mathrm{Var}[\Delta s_i^{(j)}]$ into the probability formula:

$$P\left(\Delta s_i^{(j)} > 0\right) = \Phi\left(\frac{mp(1-p)}{\sqrt{mp(1-p) + nmp^2(1-p^2)}}\right)$$

Step 6: Substitute m and n

Simplify the Expression:

$$P\left(\Delta s_i^{(j)} > 0\right) = \Phi\left(\frac{\sqrt{mp(1-p)}}{\sqrt{1 + np(1+p)}}\right)$$

Substitute $n \le \frac{2N}{k}$, $m \ge \frac{\left[\Phi^{-1}(1-\delta)\right]^2\left(1+\frac{2N}{k}p(1+p)\right)}{p(1-p)}$:

$$P\left(\Delta s_i^{(j)} > 0\right) = \Phi\left(\frac{\sqrt{mp(1-p)}}{\sqrt{1 + np(1+p)}}\right) \ge 1 - \delta$$