Multi – sets and ordered structures

Lecture 7
ICOM 4075
Multi-sets (or bags)

Definition: A multi-set, better known as “bag”, is a **collection of objects that may contain repeated occurrences of elements**

Remarks:

• Two bags A and B are equal if the number of occurrences of each element in A or B is the same in each bag

• The order in which the elements are listed does not alter the equality between bags
Multi-sets (or bags)

**Definition**: A multi-set, better known as “bag”, is a collection of objects that may contain repeated occurrences of elements.

**Remarks**:
- Two bags A and B are equal if the number of occurrences of each element is the same in each bag.
- The order in which the elements are listed does not alter the equality between bags.

Just as sets, bags are unordered structures.
Sets and bags model storing devices

- Sets and bags can be regarded as mathematical abstractions of storing spaces

---

**HIGH-LEVEL MODEL OF A PROCESS**

- Input data
- Get data item
- Do something
- Store partial result
  - Get new data item
  - Do something else
  - Store partial result
  - Etc...
- Output final result

**HIGH-LEVEL MODEL OF THE STORING DEVICE**

- Store input
- Data item 1
- Data item 2
- Data item 3
- Data item 4
- Etc...

**Input/Output (I / O)**
Sets and bags model storing devices

- Sets and bags can be regarded as mathematical abstractions of storing spaces.
- When using sets, only the first occurrence of an element is stored. When using bags all occurrences are stored.
- Thus, sets reduce storing space but lose some information (the number of occurrences). Bags use more space than sets, but contain more information.
Illustrations

Problem 1: A researcher is interested in determining whether trees in a list of 120 different species are present in an area of the Amazon forest. Since each specie has a different radiation signature, a plane equipped with instruments for detecting these signatures is set to fly over the area. The data is collected in a computer for further analysis.

Question: What would you use to model the collection of the data: a set or a bag?
Airborne data collection

Data collector connected to solar radiation sensor

Solar refraction
Answer to problem 1

The researcher is interested in determining whether trees in a list of 120 different species are present in an area of the Amazon forest. Thus, her or his data collection table will look like:

<table>
<thead>
<tr>
<th>Species 1</th>
<th>Signature detected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Species 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Signature not detected</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Species 120</td>
<td>Signature detected</td>
</tr>
</tbody>
</table>

Since repeated signatures do not count, a set is the best model for this data collection event.
Illustrations (cont.)

**Problem 2:** A grading directive issued by the Dean of Academics Affairs of Princeton’s School of Engineering sets the goal of keeping the number of grades “A” up to a maximum of 32% of all grades. A programmer is charged with the duty of reporting the degree of compliance with this policy by the different departments, semester by semester.

**Question:** What would you recommend as a model for storing the data, a set or a bag?
Answer to problem 2

The data collector is a scanner that reads the registrar electronics files identifying A, B, C, D and F’s. In this case, each occurrence of an A counts as a different data item. The programmer’s report should look like:

<table>
<thead>
<tr>
<th>Engineering Courses</th>
<th>Students enrolled</th>
<th>A’s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Course 1</td>
<td>27</td>
<td>5</td>
</tr>
<tr>
<td>....</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Course 576</td>
<td>23</td>
<td>7</td>
</tr>
</tbody>
</table>

Thus, data storage can only be modeled as a bag.
Similarities between bags and sets

• Bags are better described by listing, although they can also be described by property

• The elements in a bag are listed or described between brackets. Example: \([a, b, a, 1, 1]\) is a bag

• A bag \(B\) is said to be a sub-bag of a bag \(C\) if any element in \(B\) occurs in \(C\) and the number of occurrences of each element in \(B\) is less than or equal to the number of occurrences of the same element in \(C\)

• Example: \([a, a, b, 1]\) is a sub-bag of \([a, b, a, 1, 1]\) but \([a, a, a, 1]\) is not a sub-bag of \([a, b, a, 1, 1]\)
Similarities between bags and sets (cont.)

• The **union** of two bags \( B \) and \( C \) is the bags that contains each element in \( B \) or \( C \) repeated as many times as its largest number of occurrences in the two bags

• **Example:**

\[
[1, 1, a, b, c, a, 1, b, b] \cup [1, 2, a, c, c, b] = [1, 1, 1, a, a, b, c, c]
\]
Similarities between bags and sets (cont.)

• The **intersection** of two bags \( B \) and \( C \) is the bag formed by the elements that appear in \( B \) and \( C \), repeated as many times as its **minimal number of occurrences between the two bags**

• Example:

\[
[a, a, 1, 1, 1, c, c] \cap [a, 1, 2, 2, c, c, c] = [a, 1, c, c]
\]
The sum of two bags

• If B and C are two bags, we define the sum of B and C, denoted B + C, with the following rule:
  “If x occurs n times in B and m times in C then x occurs n + m times in B + C”

• Example:

  \[ [1, 1, a, b, a, c] + [1, 2, 2, a, b, b, c] \]
  \[ = [1, 1, 1, 2, 2, a, a, a, b, b, b, c, c] \]
Observation

• The union of two bags corresponds indeed to the taken of the maximum number of occurrences of each of the items in each bag.

• The intersection, instead, corresponds to the taken of the minimum number of occurrences of each of the items in each bag.

• The sum corresponds to the total number of occurrences of each item in the bags.

• A use of these fact in arithmetic is next.
Recalling some arithmetic

The least common multiple (lcm) of two integers $a$ and $b$ is the smallest positive integer that is a multiple both of $a$ and of $b$. Since it is a multiple, it can be divided by $a$ and $b$. If either $a$ or $b$ is 0, then lcm($a$, $b$) is defined to be zero.
Recalling some arithmetic

The **least common multiple (lcm)** of two integers \(a\) and \(b\) is the smallest positive integer that is a multiple both of \(a\) and of \(b\). Since it is a multiple, it can be divided by \(a\) and \(b\). If either \(a\) or \(b\) is 0, then \(\text{lcm}(a, b)\) is defined to be zero.

The **greatest common divisor (gcd)** of two non-zero integers is the largest positive integer that divides both numbers.
Defining a new bag

Recall that "Prime Factorization" is finding which prime numbers you need to multiply together to get the original number.

**Example**: What is the prime factorization of 12?

**Answer**: 12 = 2 × 2 × 3 = 2^2 × 3

Let’s define the bag \( p(n) = \{x : x \text{ is a prime that occurs in the prime factorization of } n\} \)

**Example**: \( p(12) = \{2, 2, 3\} \)
Defining a new bag

Recall that "Prime Factorization" is finding which prime numbers you need to multiply together to get the original number.

Example: What is the prime factorization of 12?

Answer: \(12 = 2 \times 2 \times 3 = 2^2 \times 3\)

Let’s define the bag \(p(n) = \{x : x \text{ is a prime that occurs in the prime factorization of } n\}\)

Example: \(p(12) = \{2, 2, 3\}\)

Notice that this is a bag defined by property.
Playing with bags $p(n)$

Consider the union of

$p(54) = [2, 3, 3, 3], p(12) = [2, 2, 3],
[2, 3, 3, 3] \cup [2, 2, 3] = [2, 2, 3, 3, 3]$

The number represented by the resulting bag is

$4 \times 27 = 108$. This is, $p(54) \cup p(12) = p(108)$

But notice that 108 is also the least common multiple of 54 and 12!

Is $p(n) \cup p(m) = p(lcm \ of \ n \ and \ m)$ always true?
Playing with bags p(n)

Consider now the intersection of

\[ p(54) = [2, 3, 3, 3], \quad p(12) = [2, 2, 3], \]
\[ [2, 3, 3, 3] \cap [2, 2, 3] = [2, 3] \]

The number represented by the resulting bag is

\[ 2 \times 3 = 6. \] This is, \( p(54) \cap p(12) = p(6) \)

But notice that 6 is also the greatest common divisor of 54 and 12!

Is \( p(n) \cap p(m) = p(gcd \text{ of } n \text{ and } m) \) always true?
Ordered structures

Sets and bags are oblivious to order. Next we introduce structures that do preserve the order.

**Definition**: A tuple is an ordered bag. A k-tuple is a tuple with k elements. The elements of a tuple are listed between parenthesis.

**Example**: $(1, 1, a, b, a)$ is a 5-tuple.

**Definition**: Two tuples are equal if they are equal as bags and their elements are listed in the same order.

For example, $(1, 1, a, b, a) \neq (1, a, 1, b, a)$.
Ordered structures

Sets and bags are oblivious to order. Next we introduce structures that do preserve the order.

**Definition**: A tuple is an ordered bag. A k-tuple is a tuple with k elements. The elements of a tuple are listed between parenthesis.

**Example**: \((1, 1, a, b, a)\) is a 5-tuple.

**Definition**: Two tuples are equal if they are equal as bags and their elements are listed in the same order.

For example, \((1, 1, a, b, a) \neq (1, a, 1, b, a)\)
Tuples as models of storing devices

Tuples contain more information than bags. Just as bags, tuples store all occurrences of an element, but in addition, they store the order in which these occurrences happen.

Problem 3: An atmospheric scientist is interested in correlating the average temperature and atmospheric pressure oscillations by hour and the occurrences of heavy rain events in the tropical forest. A device that captures both temperature and atmospheric pressure, and transmit them to a computer center via satellite has been installed in the forest.

Question: How would you model the storing device?
Answer to problem 3

The scientist expects to fill a table of data of the form of

<table>
<thead>
<tr>
<th>Time</th>
<th>Pressure</th>
<th>Temperature</th>
<th>Rain</th>
</tr>
</thead>
<tbody>
<tr>
<td>00:00</td>
<td>1013.25 mbs</td>
<td>65 F</td>
<td>0 mms</td>
</tr>
<tr>
<td>01:00</td>
<td>1013.32 mbs</td>
<td>66 F</td>
<td>0 mms</td>
</tr>
<tr>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
<tr>
<td>23:00</td>
<td>1012.98 mbs</td>
<td>63 F</td>
<td>0.65 mms</td>
</tr>
</tbody>
</table>

The order in which the measurements are taken is of the essence. So, the right model is a tuple
Definition: The **Cartesian product** of two sets $A$ and $B$, is defined to be the set of all 2-tuples formed with one element in $A$ and one element in $B$. The Cartesian product is denoted $A \times B$

Example: Let $A = \{a, b, c\}$, $B = \{1, 2\}$. Then,

$$A \times B = \{(a, 1), (a, 2), (b, 1), (b, 2), (c, 1), (c, 2)\}$$

Formally: $A \times B = \{(x, y): x \text{ in } A \text{ and } y \text{ in } B\}$
Cartesian product of $n$ sets

- The previous definition is naturally extended to a collection of $n$ sets $A_1, A_2, \ldots, A_n$ as follows:

$$A_1 \times A_2 \times \cdots \times A_n = \{(x_1, x_2, \ldots, x_n) : x_i \in A_i, i = 1, \ldots, n\}$$

This is, the set of all $n$-tuples with one element in each set of the collection, taken in the order in which the collection is listed.
More on Cartesian products

A couple of properties:

• $A \times B \neq B \times A$ (in general. There are exceptions)
• $A \times \emptyset = \emptyset \times A = \emptyset$

An additional definition: The null tuple is defined to be the tuple with no elements. The null tuple is denoted $(\ )$.

Remark: The null tuple is not the same as the empty set !!!
A special set of Cartesian products

Definition: Let $A$ be a set. Then, we define “$A$ to the $n$-th” power as:

$A^0 = \{(())\}$

$A^1 = A$

$A^2 = A \times A$

$A^3 = A^2 \times A = A \times A^2$

$\vdots$

$A^n = A \times A^{n-1} = A^{n-1} \times A$, for $n \geq 1$. 
Remarks

• The n-th power of a set A is not the power set of A
• The operation of taking the n-th power of a set models data aggregation. The following example illustrates this:

Example: Let A = \{x: x is a temperature in Fahrenheit scale\}

- ( ) ∈ A^0 (no measurements yet)
- (56) ∈ A^1 (first measurement recorded)
- (56, 59) ∈ A^2 (second measurement recorded)
- (56, 59, 64) ∈ A^3 (third measurement recorded)

The tuple with all the temperatures per hour in a whole day belongs to A^{24}
Relations: another interpretation of tuples

**Definition:** An **n-ary relation** over a finite collection of sets \( A_1, \ldots, A_n \) is a **statement that describes a subset of** \( A_1 \times \cdots \times A_n \)

**Example:**

Let \( A_1 = \{1, 2, 3\} \), \( A_2 = \{-1, 2\} \), and \( A_3 = \{0, 3\} \).

Define the property:

\[ P : "(x, y, z) \text{ if and only if } x \leq y \text{ and } y < z". \]

This statement defines the subset of \( A_1 \times A_2 \times A_3 \):

\[ P = \{(1, 2, 3), (2, 2, 3)\} \]

which is called relation \( P \).
Further notes on relations

• As shown in the previous example, it is customary to use the same symbol (P in the previous example) to denote the statement that defines the relation and the set of tuples in the relation.

• A relation may be false in some instances.

Example: Let P: “X was introduced to Y by Z”. The corresponding set of 3-tuples is:

\[ P = \{(X, Y, Z): X, Y \text{ and } Z \text{ are people that satisfy } P\} \]

A choice of elements such as:

\[ X = \text{Napoleon}, \ Y = \text{Plato}, \ Z = \text{Newton} \text{ makes } P \text{ false. This is, } (X, Y, Z) \text{ does not belong to } P. \]
A numeric example

Consider the relation D over the set of all pairs (2 – tuples) of nonnegative real numbers defined as D: “(x, y) if and only if \( \sqrt{(x - 2)^2 + (y - 3)^2} \leq 1 \).”

As a set, D corresponds to the circle of radio 1:

The 2 – tuples in this relation are all in the shaded area.
Summary

• Concept of multi-set or bag
• Abstract models of data storage: sets vs. bags
  – Problems: Would you recommend a set or a bag?
• Sub-bags, union, intersection and summation of two bags
• Ordered structures: tuples
• Cartesian products and tuples. Properties.
  – The null tuple; the powers of a set
• Relations as subsets of tuples
  – Relational databases
Exercises

1. Define an empty bag as a bag with no elements, and use logic to decide whether the following statements are true or false:
   – “The union of a bag A and the empty bag is the bag A”
   – “The intersection of a bag and the empty bag is the empty bag”
   – “The summation of a bag A and the empty bag is the empty bag”

2. Find the union, intersection, and sum of the following bags
   – [a, a, b] and [a, c, c, b, b, b]
   – [1, 1, 1, 2, 3, 4] and [1, x, y, 4, 4]
   – [x, x, [a, a], [a, a]] and [a, a, x, x]
   – [a, a, [b, b], [a, [b]]] and [a, a, [b], [b]]

3. For each equation, find a bag that satisfies it
   – B + [2, 2, x, 4, 4] = [2, 2, x, y, 4, 4, 4]
   – B ∩ [3, 4, 5, 5] = [4, 5]
   – B ∪ [2, 2, 3, 3, 4, 4] = [2, 2, 3, 3, 3, 3, 4, 4, 4, 4, x, y, y, z]
Exercises (cont.)

4. Let A be a sub-bag of a bag B. Define the complement of A with respect to B and verify which properties of set complement are preserved by your definition of bag complement.

5. Use the definition in your answer to problem 5 to define a difference between bags.

6. Draw a graph of the relation over the set of integers R: “(x, y) if and only if max |x – 5| ≤ 4 and max |y – 3| ≤ 2”

7. Draw a graph of the relation over the set of integers R: “(x, y) if and only if max |x – 5| ≤ 4 or max |y – 3| ≤ 2”

8. Draw a graph of the relation over the set of integers R: “(x, y) if and only if max (|x – 5| + |y – 3|) ≤ 2”
9. Let \( n \) and \( m \) be integers. Consider the bags \( p(n) \) and \( p(m) \). How would you characterized the bag \( p(n) + p(m) \)?

10. Propose an adequate model for storing: A manufacturing plant produces 12 different items. A production manager is interested in collecting data to detect some potential inefficiencies in the production lines. In particular, the manager wants to know

- Which of the 12 production lines produces the highest number of items per day and which the least
- What is the percentage of defective items in each line, per day

11. Propose an adequate model for storing: A sales manager of the same plant wants to conduct surveys on customers preferences. The sales manager wants to collect data from a two-day survey of customers in 20 supermarkets. The data should determine how many of the 12 items (if any) are actually purchased by each surveyed customer.