

Summary Module 7

1. Data dictionary is a catalogue of all data used in an application. Each entry in the dictionary has the name of the data, its type, origin, destination, range of values, length and where used.
2. It is useful as a documentation and assists in maintaining systems and also in designing data input.
3. Data dictionary is developed using data flows in a DFD and contents of data stores in DFD
4. Data may be input to a computer off-line or on-line. In on-line entry a user enters data interactively via a video terminal connected to the computer. In off-line data entry data filled in forms are entered by operators in a separate PC or a data entry machine.
5. Off-line data entry is suitable if the number of records is very large.
6. In off-line data entry, batches of data are formed and entered. They are checked by a validation program and the corrected records are stored in a file.
7. To reduce errors in input, it is essential to carefully design the forms used for entering data.
8. For interactive data input, special screens are designed on video terminals for easy data entry. Errors in data entry are instantly detected by a validation program during data entry and can be immediately corrected.
9. Common methods of interactive data input is by use of menus, templates and interactive commands for data entry.

10. A menu method is used to pick one out of many alternatives, a template method to enter new data, and a command method to add and delete data.
11. These methods are combined to provide a user the most appropriate technique for a particular type of interactive data entry.
12. Important data elements are coded. Codes are necessary for unique identification, easily cross-referencing and efficient storage and retrieval.
13. There are many methods for coding. An ideal code must be concise, expandable, meaningful, comprehensive and precise. It is not possible to incorporate all these ideal features in a code.
14. Codes are classified as: (i) Serial number codes, (ii) Block codes, (iii) Group classification codes, and (iv) Significant codes. Group classification codes and Significant codes are most meaningful, expandable, precise and comprehensive. They are, however, not concise, Serial and Block number codes are more concise. They are also precise and expandable but are not meaningful and comprehensive.
15. Any error made in entering important data fields such as account codes and identification codes must be detected during data entry.
16. The most common errors made during data entry are: a single digit is incorrectly entered or any two digits in the code are interchanged. These errors are called respectively single transcription and transposition errors and account for 96% of all data entry errors.
17. Given a code, the digits in it starting from the last digit are multiplied by weights 2, 3, 4, etc., and the products are added. The sum is divided by 11. The remainder is subtracted from 11. This number (which is called a check digit) is

- appended as the last digit of the code. The code constructed in this way is called a modulus-11 check digit code.
18. After data entry the digits in the code starting from the last digit are multiplied by weights 1, 2, 3, 4, etc., and the products are added. The sum is divided by 11. If the remainder is not zero then there is an error in the code.
 19. Modulus-11 check digit code guarantees detection of all single transcription and transposition errors. It also detects 95% of all other errors.
 20. It is essential to design good data validation programs to prevent data entry errors from corrupting files of input data. Validation programs need information for detecting errors. This information is provided by controls exercised during data preparation.
 21. Important control mechanisms are; giving unique sequence numbers to each data record, providing a batch control record containing a count of number of records and a total of one of the fields..
 22. The same data is entered by two different persons and compared to reduce transcription errors.
 23. Besides this, individual data fields are checked using information on their range of allowed values, range of reasonable values, and relationships between different fields. Batch control provides information to detect incorrect values entered, missing records, and data in the wrong sequence.
 24. With the advent of Personal Computers, remote terminals connected to a computer and local computer networks, considerable amount of data is entered in files interactively.

Question Bank – 7

- 7.1 What is a data dictionary?
- 7.2 Why is a data dictionary necessary?
- 7.3 What are the main advantages of creating a data dictionary?
- 7.4 What data about a data element is stored in a data dictionary?
- 7.5 For the requirements statement given in PPT 7.1.3 (Vendor supplying items to a company) develop the data dictionary entry for vendor code, vendor name and vendor address.
- 7.6 What is the difference between on-line and off-line data entry?
- 7.7 Why are input data records divided into batches for off-line data entry?
- 7.8 What is the purpose of a data validation program?
- 7.9 What are the main principles used in designing forms for data entry?
- 7.10 A good and a bad design for entering date in a form is given in Section PPT 7.2.5 and 7.2.6. What are the reasons for saying that one of them is good and the other bad?
- 7.11 Design a form to be used by a salesman to report to the office about the sales executed by him at different customer locations.
- 7.12 What is the main difference between menus, templates and command modes of interactive data entry? When is each of these modes appropriate?
- 7.13 Design a dialogue hierarchy for entering data on customers (of a manufacturer).
- 7.14 Design a dialogue hierarchy and the screens for a system used to reserve seats in long distance buses.
- 7.15 Why are data fields coded in an information system?

- 7.16 Can the name of a person be used as a code, for say, his bank account? If not, why?
- 7.17 What are the requirements of a good coding scheme?
- 7.18 Is a concise code comprehensive? If not, why?
- 7.19 Is a meaningful code necessarily comprehensive?
- 7.20 Is a comprehensive code necessarily meaningful?
- 7.21 Is a precise code necessarily concise?
- 7.22 What is the advantage of a serial number code? Why is it not normally used?
- 7.23 What is the main advantage of block codes?
- 7.24 Design a group classification code to code (i) motor vehicles,, (ii) music cassettes, and (iii) books
- 7.25 Is a group classification code meaningful?
- 7.26 Give an example of a significant code. Are significant codes expandable?
- 7.27 Add a Modulus-11 check digit to the codes (i) 48467, (ii) 96432, and (iii) 87646257.
- 7.28 Modulus-37 check is suitable for alphanumeric codes. Add a modulus-37 character to the codes (i) 4AB9W, (ii) XBY483, and (iii) CAZ4642.
- 7.29 The following code was entered by an operator:449632. The last digit is a modulus-11 check digit. Is this code correct?
- 7.30 If a code uses hexadecimal digits, what should be N if the modulus- N check digit system is to be used with such codes? What are the allowable weights if single transcription and transposition errors are to be detected?

7.31 If modulus-11 check digit system is to generate detection of multiple identical digit transcription error (i.e., a code such as 45565 is wrongly entered as 48868), what should be the constraints on the weights?

7.32 A see-saw error is one in which one digit of the code is increased by x and another decreased by x . For example, 486732 becoming 456762. When can modulus N check detect such errors?

7.33 Why is it useful to assign sequence numbers for data records? What are the types of errors detected by sequence numbering?

7.34 What is the purpose of batch control record? What is the type of information contained in a batch control record?

A set of data records for student examination results has the following format:

Roll no. Name Marks (out of 100)

Paper 1 Paper 2 Paper 3 Paper 4

Design for these records a batch control record and a record control field and any other appropriate checks for the fields.

7.35 Give some example of fields for which a radix error check is appropriate.

7.36 What is the difference between range check and a radix check?

7.37 What are the appropriate range checks for the age of individuals in an employee file, a high school student file, and height of students in a student file.

7.38 Give some examples of fields where reasonableness check would be applicable.

7.39 Give some examples of inter-field relationship checks.

7.40 What is the main difference between menus, templates and command modes of interactive data entry? When is each of these modes appropriate?

7.41 Design a dialogue hierarchy for entering data on customers (of a manufacturer).

7.42 Design a dialogue hierarchy and the screens for a system used to reserve seats in long distance buses.

References

1. Most of the material in this module has been adapted from the book “Analysis and Design of Information Systems”, 2nd Edition, by V.Rajaraman, Prentice Hall of India, 2003. Chapter 5 (pp. 49-52) and Chapter 11 (pp.154-170).
2. Good material on Data Dictionary is found in K.E.Kendall and J.E.Kendall , “Systems Analysis and Design”, 5th Edition, Pearson Education Asia, 2003. Chapter 10 on Data Dictionaries. Chapter 16 Designing Effective Input.