Data Structures

Wen-Chih Peng (彭文志)
Department of Computer Science
National Chiao Tung University
Fall, 2012
Course Information

- Instructor: Wen-Chih Peng (彭文志)
- Office: EC 542 (Tel: 31478)
- E-mail: wcpeng@cs.nctu.edu.tw
- TA: 廖忠訓 (EC621)
- Web site (e3.nctu.edu.tw)
Course Information (Cont’d)

- **Prerequisite**
  - object oriented programming C++

- **Grading (Tentative)**
  - Midterm 30%, final 30% and 4-5 homeworks (40%)
Course materials

- Introduction (Concepts, Recursion, Algorithm Analysis)
- Arrays
- Stacks & Queues
- Lists
- Trees
- Graphs
- Sorting
- Hashing
- Advanced Data Structures
Importance

- Basis of all fields in computer science
- Data structures + Algorithms + Coding skills = Programming
Web search engine

1. Web crawling
2. Indexing
3. Searching
Big-Table for Storing Web pages

圖一、以儲存 Web Pages 為例
Problem:
- Given a document and a string, find the all appearances of the string in the document.

Related techniques:
- String Matching (Brute-Force, KMP)
Tiger Woods Clinches PGA Player of the Year Award
2006/9/8
By Michael Buteau NEW YORK, Bloomberg
Tiger Woods, who has won his past five tournaments and seven overall this season, clinched his eighth PGA Player of the Year Award.
Keywords as an Indexing

- **Problem:**
  - Given a document and a list of stop words, find all words and their frequencies.

- **Related techniques:**
  - Array
  - Linked list
  - Binary search tree
Ranking Web pages
Ranking Web Pages

- **Problem:**
  - There are many web pages containing the search keyword. But, how to score them?

- **Algorithms: PageRank or HITS:**
  - Graph

![Graph diagram](image)
How to program

- Requirements
- Analysis: **bottom-up vs. top-down**
- Design: data objects and operations
- Refinement and Coding
- Verification
  - Program Proving
  - Testing
  - Debugging
Algorithm

- **Definition**
  An *algorithm* is a finite set of instructions that accomplishes a particular task

- **Criteria**
  - input
  - output
  - definiteness: clear and unambiguous
  - finiteness: terminate after a finite number of steps
  - effectiveness: instruction is basic enough to be carried out
Specification vs. Implementation

- **Operation specification**
  - function name
  - the types of arguments
  - the types of the results

- **Implementation**
  - Program language (e.g., C++, Java)
  - How to implement using one of data structures learned
Evaluation (Measurement)

**Criteria**
- Is it correct?
- Is it readable?

**Performance Analysis** (machine independent)
- space complexity: storage requirement
- time complexity: computing time

**Performance Measurement** (machine dependent)