Experiment in Compiler Construction
Indexer Design

School of Infomation and Communication Technology

Hanoi University of Science and Technology

Tasks of an indexer

- Reading text from IO file.
- Creating an index table for each words in text – a stream of characters
- Index table must be
  - Containing every words in text
  - # of its appearances
  - Its position inside text, i.e. # of line

Task detailed

- Textual IO file : "vanban.txt" contains a journal paragraph talking about a suicide bombing in the Golf War.

- A word is determined by a sequence of continuous characters without *space* or *tab* character inside.

- Every indexed word must be in lower-case letter.

Task detailed (cont.)

- Index table does not contain any stop-words that are listed in "stopw.txt".

- Index table is able to arrange its words in alphabetical order.

Task detailed (cont.)

- For example

answer 7, 8, 12, 15

ant 2,4,6

…

  - # of appearances
  - # of lines

Program

- Reading a word
- Checking its validity
- Saving this word into an index table
- Updating its statistics, i.e. # of appearances, # of lines

Programming skills

- Reading and writing IO file
- Manipulating strings and ASCII characters
  - To compare two strings
  - To identify special characters such as space, tab, new line etc…
- Wrapping gcc program by Makefile