



ĐẠI HỌC BÁCH KHOA HÀ NỘI
VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

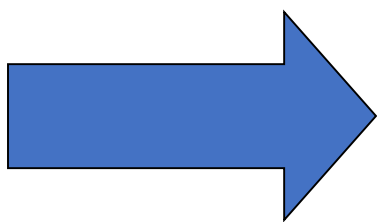
XÂY DỰNG CHƯƠNG TRÌNH DỊCH

Môn học sẽ nghiên cứu

- Cách thức làm việc của bộ xử lý ngôn ngữ và chương trình dịch
- Sinh mã máy cho những cấu trúc ngôn ngữ cụ thể
- Thiết kế ngôn ngữ: Cú pháp và ngữ nghĩa

Tại sao cần nghiên cứu CT dịch?

- Rèn kỹ năng phát triển ứng dụng quy mô lớn
- Làm việc với các cấu trúc dữ liệu phức tạp
- Tìm hiểu sự tương tác giữa các giải thuật



Bước chuẩn bị cho những dự án lớn trong tương lai.

Những vấn đề chính

- Bộ xử lý ngôn ngữ
- Cấu trúc của một trình biên dịch (1 pha)
- Biểu diễn cú pháp: văn phạm hình thức, BNF và sơ đồ cú pháp
- Phân tích từ vựng
- Phân tích cú pháp: quay lui và tiền định
- Văn phạm LL(k) và phân tích kiểu đệ quy trên xuống
- Ngôn ngữ lập trình KPL: cú pháp và ngữ nghĩa
- Phân tích ngữ nghĩa
- Sinh mã: sinh mã trung gian và sinh mã đích
- Tối ưu mã

Tài liệu tham khảo

- Aho.A.V, Sethi.R., Ullman.J.D.
Compiler : Principles, Techniques and Tools.
Addison Wesley.1986
- Bal.H. E.
Modern Compiler Design.
John Wiley & Sons Inc (2000)
- William Allan Wulf.
The Design of an Optimizing Compiler
Elsevier Science Ltd (1980)
- Charles N. Fischer.
Crafting a Compiler
Benjamin-Cummings Pub Co (1987)

Tài liệu tham khảo (tiếp)

- Niklaus Wirth
Compiler Construction.
Addison Westley. 1996
- Andrew.W.Appel
Modern Compiler Implementation in Java
Princeton University.1998
- Nguyễn Văn Ba
Giáo trình kỹ thuật biên dịch
Đại học Bách Khoa Hà Nội.1994
- Vũ Lục
Phân tích cú pháp
Đại học Bách Khoa Hà Nội.1990
- Bài giảng về ngôn ngữ và phương pháp dịch
- www.sourceforge.net

Đánh giá kết quả học tập

- Giữa kỳ (30%): Tự luận – Lập trình phân tích từ vựng và phân tích cú pháp
- Cuối kỳ (70%): Thi trắc nghiệm



ĐẠI HỌC BÁCH KHOA HÀ NỘI
VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

Bài 1.

Bộ xử lý ngôn ngữ và trình biên dịch

Ngôn ngữ lập trình cấp cao

- Các ngôn ngữ lập trình được chia thành 5 thế hệ.
- Việc phân chia cấp cao hay thấp phụ thuộc mức độ trừu tượng của ngôn ngữ
 - Cấp thấp : gần với máy
 - Cấp cao : gần với ngôn ngữ tự nhiên

Ngôn ngữ lập trình thế hệ thứ nhất và thứ hai

- Thế hệ thứ nhất : ngôn ngữ máy
- Thế hệ thứ hai : Assembly
- Các ngôn ngữ thuộc thế hệ thứ nhất và thứ hai là ngôn ngữ lập trình cấp thấp

Ngôn ngữ lập trình thế hệ thứ ba

- Dễ hiểu hơn
- Cho phép thực hiện các khai báo, chẳng hạn biến
- Phần lớn các ngôn ngữ cho phép lập trình cấu trúc
- Ví dụ: Fortran, Cobol, C, C++, Basic

Ngôn ngữ lập trình thể hệ thứ tư

- Thường được sử dụng trong một lĩnh vực cụ thể (chẳng hạn thương mại)
- Dễ lập trình, xây dựng phần mềm
- Có thể kèm công cụ tạo form, báo cáo
- Ví dụ :SQL, Visual Basic, Oracle (SQL plus, Oracle Form, Oracle Report). . . .

Ngôn ngữ lập trình thể hệ thứ năm

- Giải quyết bài toán dựa trên các ràng buộc đưa ra cho chương trình chứ không phải giải thuật của người lập trình.
- Việc giải quyết bài toán do máy tính thực hiện
- Phần lớn các ngôn ngữ dùng để lập trình logic, giải quyết các bài toán trong lĩnh vực trí tuệ nhân tạo

Đặc trưng của ngôn ngữ lập trình cấp cao

- Độc lập với máy tính
- Gần với ngôn ngữ tự nhiên
- Chương trình dễ đọc, viết và bảo trì
- Muốn thực hiện chương trình phải dịch sang ngôn ngữ máy
- Chương trình thực hiện chậm hơn

Cú pháp và ngữ nghĩa của ngôn ngữ lập trình

- Cú pháp : Chính tả và văn phạm của các cấu trúc ngôn ngữ
- Ngữ nghĩa : Ý nghĩa và hiệu quả của các cấu trúc ngôn ngữ

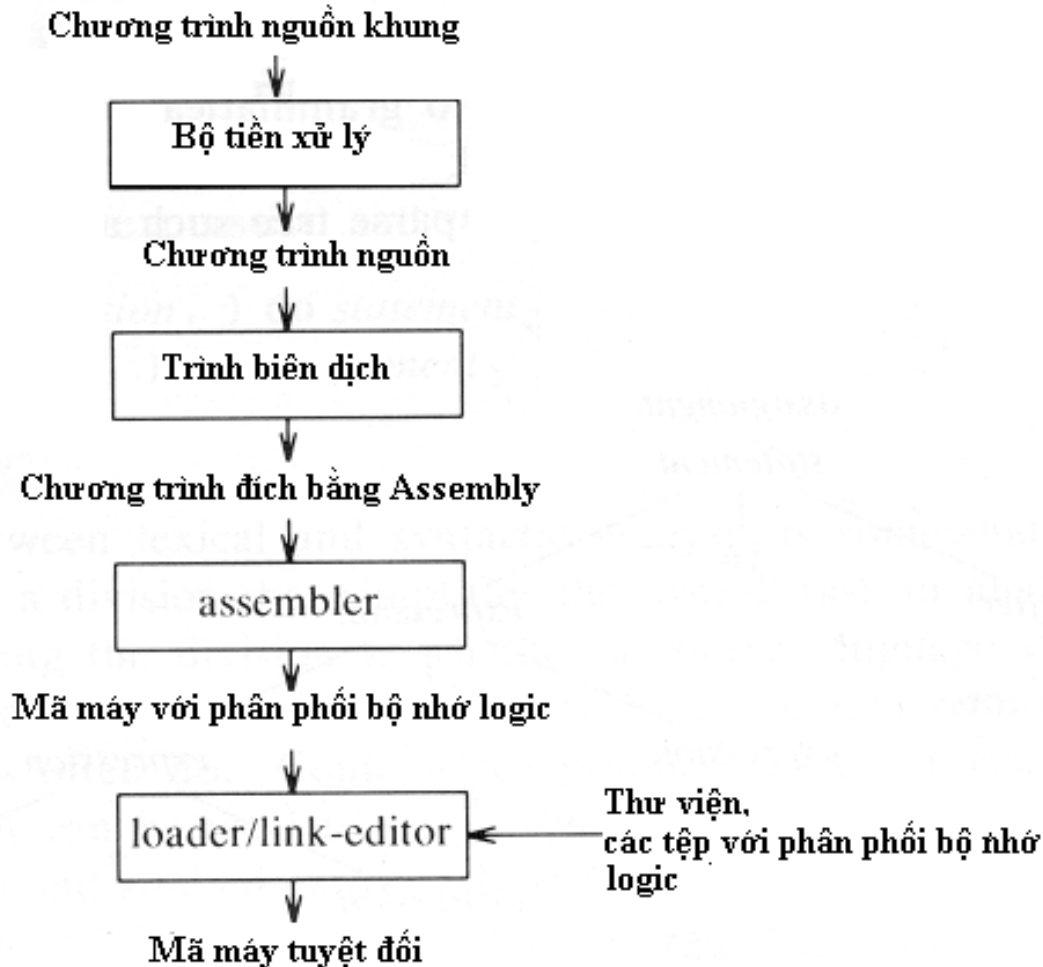
Bộ xử lý ngôn ngữ (Language Processor)

- Phần mềm đọc một chương trình viết bằng một ngôn ngữ (ngôn ngữ nguồn) dịch sang một chương trình tương đương trên ngôn ngữ khác (ngôn ngữ đích)
- Nếu chương trình đích viết trên mã máy, chương trình có thể được người dùng ra lệnh thực hiện

Ví dụ về bộ xử lý ngôn ngữ

- **Compiler: dịch từ ngôn ngữ nguồn sang mã máy**
- **Assembler: dịch từ Assembly sang mã máy**
- **Interpreter: dịch và thực thi trực tiếp từng thao tác của chương trình nguồn dựa trên dữ liệu do người sử dụng cung cấp**
- **Compiler – Compiler: bộ sinh compiler**

Bộ xử lý ngôn ngữ cho những ngôn ngữ cho phép viết chương trình trên nhiều tệp

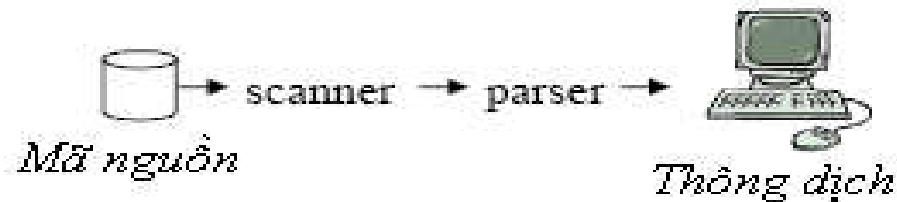


Compiler & Interpreter

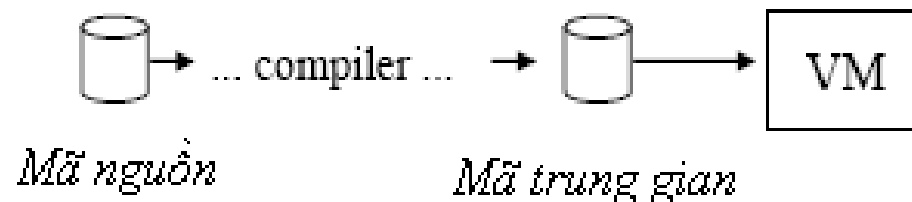
- Compiler : Dịch trực tiếp ra mã máy



- Interpreter : Trực tiếp thực hiện từng lệnh mã nguồn



- Biến thể của Interpreter : thông dịch mã trung gian



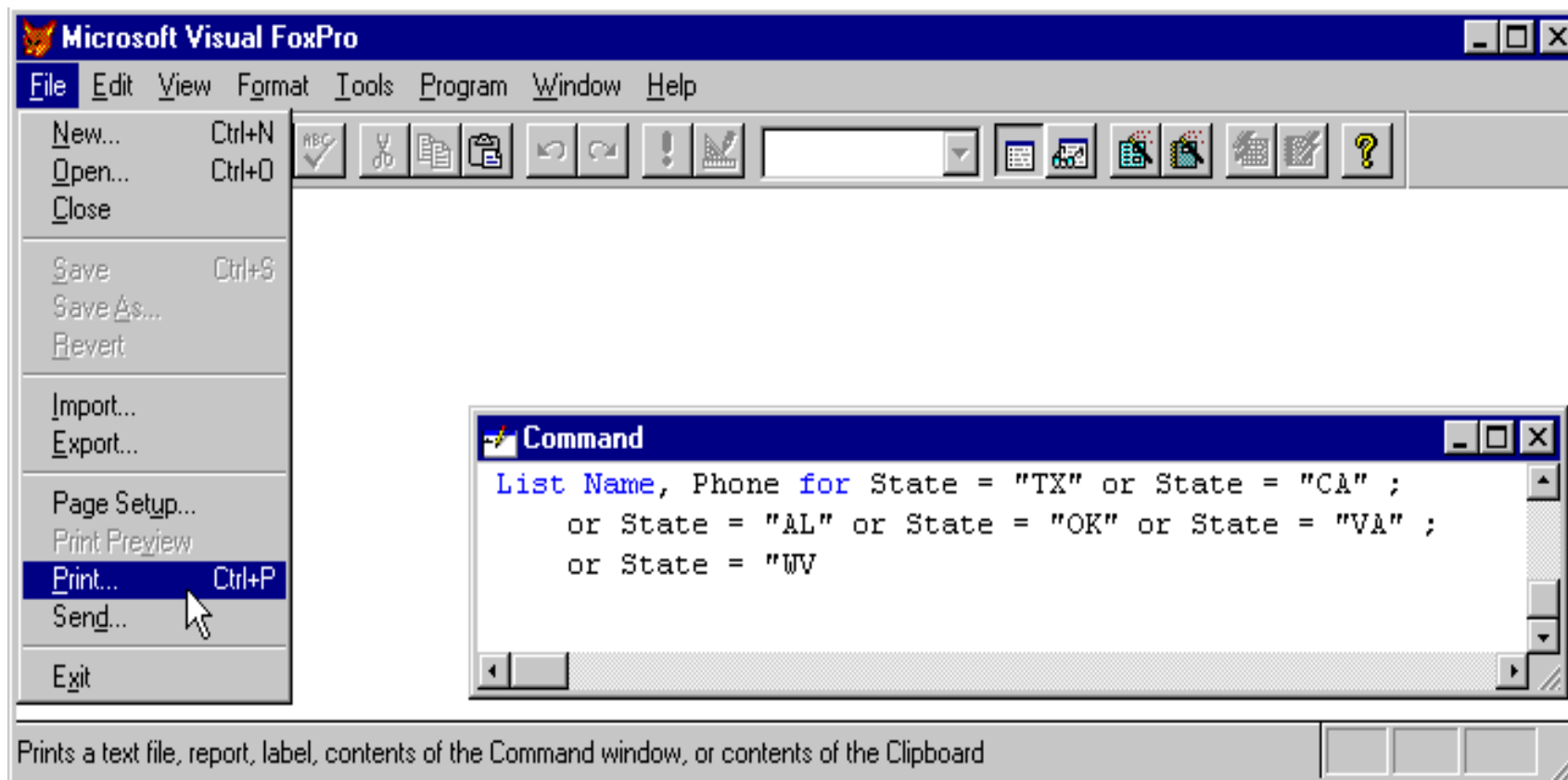
Interpreter: Trình thông dịch

- Một số ngôn ngữ sử dụng trình thông dịch cho phép dịch và chạy trực tiếp từng lệnh
- Mỗi lệnh được dịch thành một đoạn chương trình trong một ngôn ngữ trung gian. Ngôn ngữ trung gian dùng trình dịch compiler.
- Ngôn ngữ hoàn toàn dùng trình thông dịch: Foxpro
- Ngôn ngữ kết hợp thông dịch và biên dịch: Visual Basic, Python

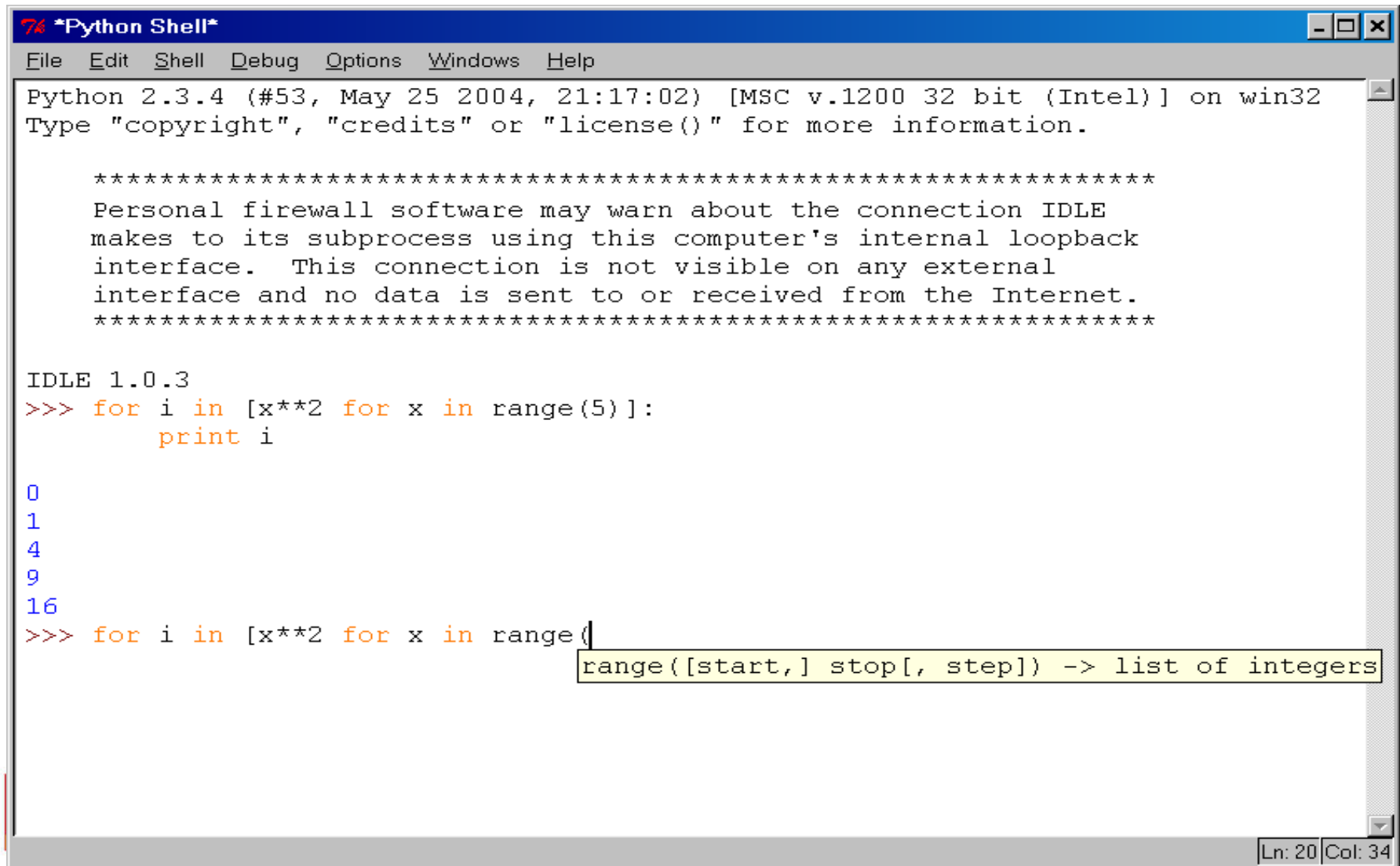
Ngôn ngữ dùng interpreter: Foxpro

- Hai chế độ làm việc
 - Cửa sổ lệnh: Thực hiện từng lệnh
 - Chương trình: Chạy từng lệnh. Các lệnh trước lệnh đầu tiên bị lỗi trong chương trình vẫn được thực hiện

Cửa sổ lệnh của Foxpro



Thực hiện từng lệnh trên Python



The screenshot shows a Python Shell window titled "Python Shell" with a menu bar (File, Edit, Shell, Debug, Options, Windows, Help). The main text area displays the following content:

```
Python 2.3.4 (#53, May 25 2004, 21:17:02) [MSC v.1200 32 bit (Intel)] on win32
Type "copyright", "credits" or "license()" for more information.

*****
Personal firewall software may warn about the connection IDLE
makes to its subprocess using this computer's internal loopback
interface. This connection is not visible on any external
interface and no data is sent to or received from the Internet.
*****

IDLE 1.0.3
>>> for i in [x**2 for x in range(5)]:
    print i

0
1
4
9
16
>>> for i in [x**2 for x in range(
```

A tooltip is visible over the opening parenthesis of the second range function, displaying the text: `range([start,] stop[, step]) -> list of integers`.

The status bar at the bottom right indicates "Ln: 20 Col: 34".

Compiler (trình biên dịch)

- Mục đích : Dịch chương trình từ ngôn ngữ cấp cao (ngôn ngữ nguồn) sang ngôn ngữ cấp thấp (ngôn ngữ đích).
- Bản thân compiler được viết trên một ngôn ngữ gọi là ngôn ngữ thực hiện

Các công cụ liên quan đến trình biên dịch

- Trình thông dịch (Interpreter)
- Assembler
- Linker
- Loader
- Bộ tiền xử lý (Preprocessor)
- Editor
- Debugger
- Profiler