



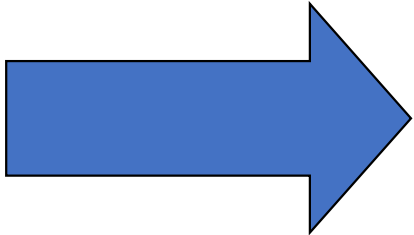
XÂY DỰNG CHƯƠNG TRÌNH DỊCH

ONE LOVE. ONE FUTURE.

- 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?

- Hiểu biết về các đặc trưng của ngôn ngữ lập trình (cú pháp, ngữ nghĩa, hướng tiếp cận)
- Rèn kỹ năng phát triển ứng dụng quy mô lớn như một trình biên dịch
 - Làm việc với các cấu trúc dữ liệu phức tạp
 - Tìm hiểu sự tương tác giữa các giải thuật
- Xây dựng những ngôn ngữ phục vụ mục đích riêng của công ty, tổ chức, ví dụ Erlang của Ericsson
- Tích hợp trình dịch vào trong các công cụ khác: Moodle, ChatGPT code
- Hiểu biết về cách tiếp cận hình thức dùng trong thiết kế phần mềm.



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

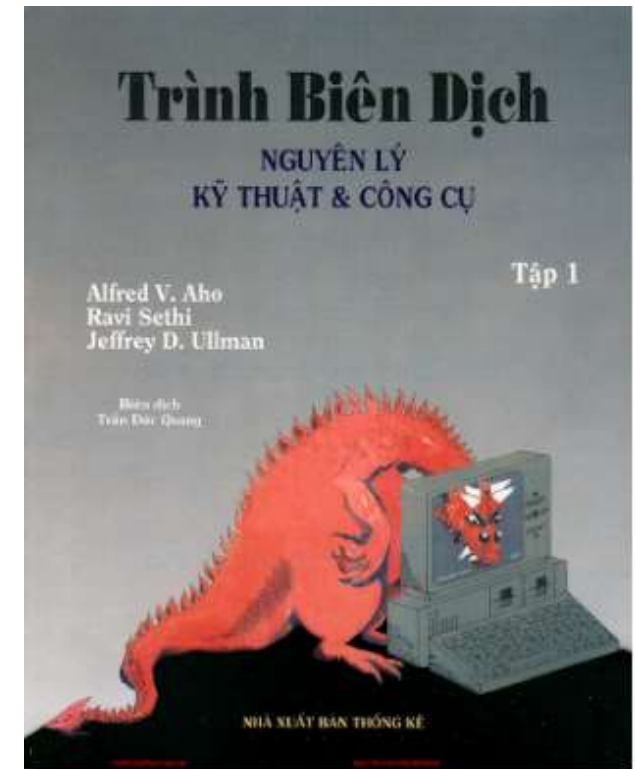
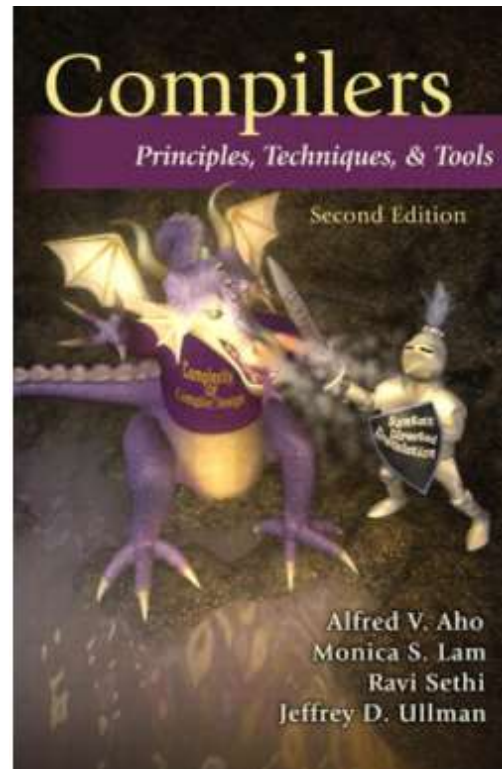
Những vấn đề chính

- Bộ xử lý ngôn ngữ
- Cấu trúc của một trình biên dịch (1 giai đoạn)
- 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, Lam M.S., Sethi.R., Ullman.J.D.
Compiler : Principles, Techniques and Tools.
Addison Wesley.2006
- Andrew.W.Appel
Modern Compiler Implementation in Java (C)
Princeton University. 2002



- Giữa kỳ (50%): Đánh giá bài thực hành. Điểm thưởng giữa kỳ do giáo viên dạy thực hành quyết định
- Cuối kỳ (50%): Thi trắc nghiệm kết hợp với tự luận.



Bài 1.

Bộ xử lý ngôn ngữ và trình biên dịch

ONE LOVE. ONE FUTURE.

- 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 thể hệ thứ nhất và thứ hai

- Thể hệ thứ nhất : ngôn ngữ máy
- Thể hệ thứ hai : Assembly
- Các ngôn ngữ thuộc thể 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



- 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). . . .

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

- 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

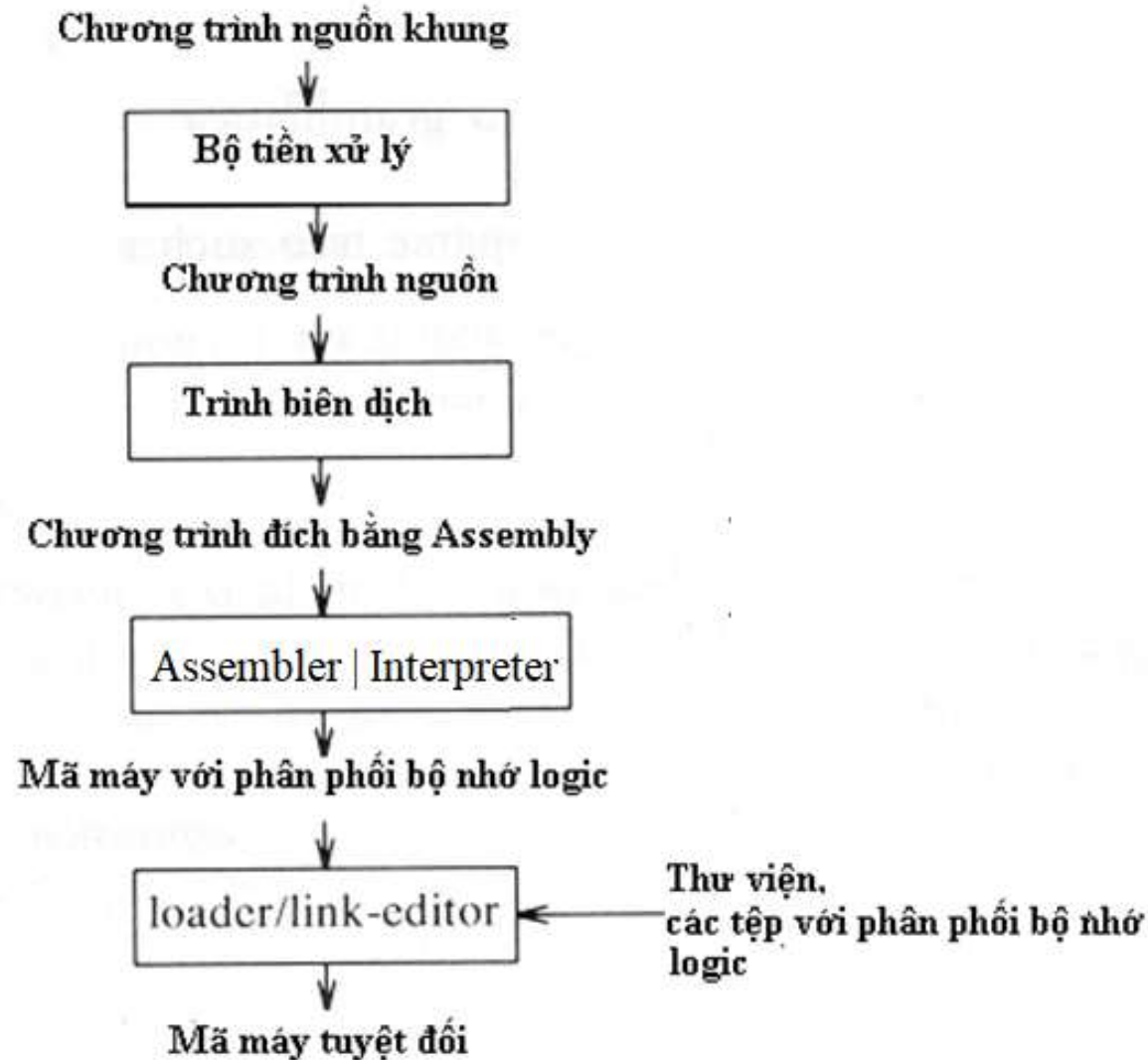
Ví dụ về bộ xử lý ngôn ngữ

- **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

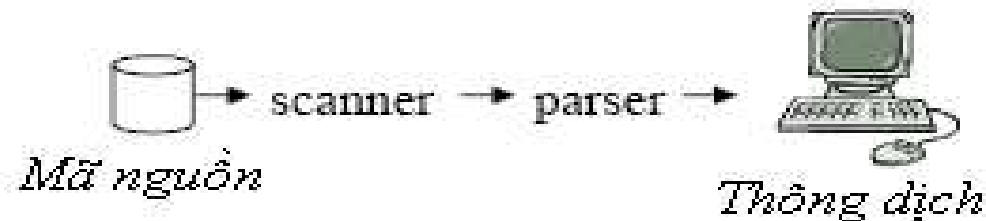
- Ngôn ngữ cho phép viết chương trình trên nhiều tệp



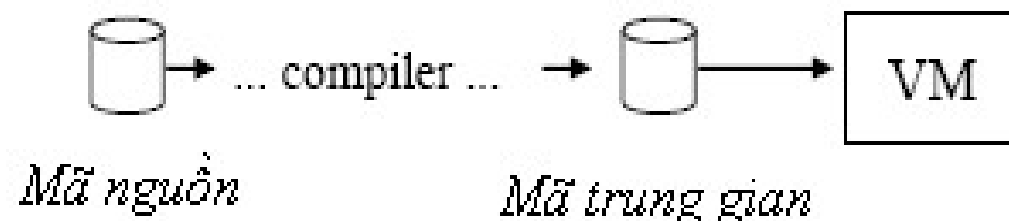
- Compiler : Dịch trực tiếp ra mã máy



- Interpreter : Trực tiếp thực hiện từng lệnh mã nguồn



- Biến thể của Interpreter : thông dịch mã trung gian



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

Ngôn ngữ biên dịch: Ngôn ngữ C

D:\Example1.cpp - [Executing] - Dev-C++ 5.11

File Edit Search View Project Execute Tools A

(globals)



Example1.cpp

```
1 #include <stdio.h>
2 main()
3 {
4     printf("Hello World");
5 }
```

Select D:\Example1.exe

Hello World


This PC > New Volume (D:)

Name	Date modified	Type
 Example1	22-Sep-21 2:57 PM	C++ So
 Example1	22-Sep-21 2:57 PM	Applicat

- 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 thông dịch: JavaScript

- JavaScript là một ngôn ngữ thuần thông dịch
- Không cần biên dịch cả chương trình mà có thể viết và thực hiện từng lệnh trên cửa sổ console.
- Tương đương chương trình Hello world, trên
- Hàm `console.log` in “Hello world” lên màn hình.

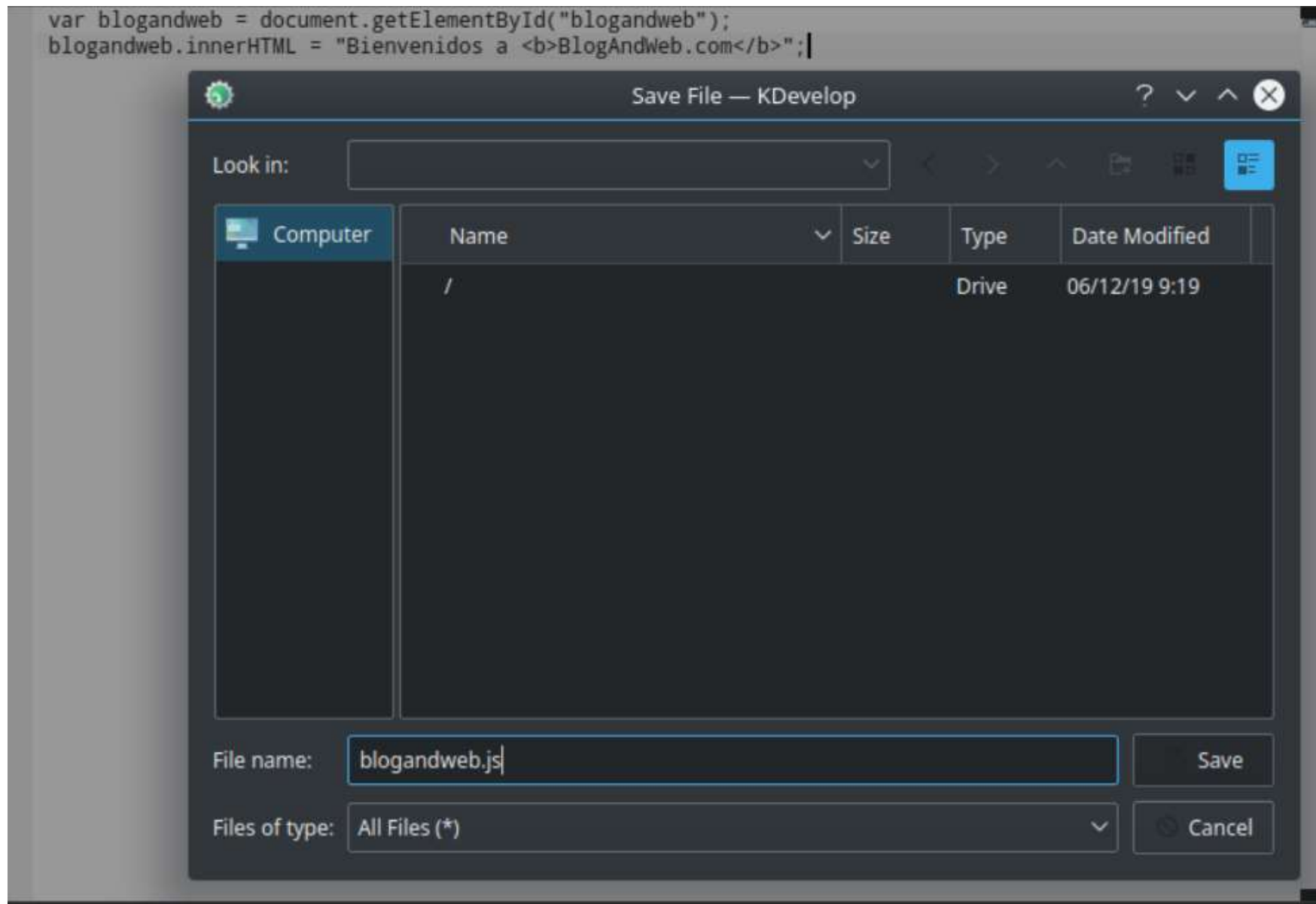


The screenshot shows a web browser's developer console with the 'Console' tab selected. The console displays the following sequence of events:

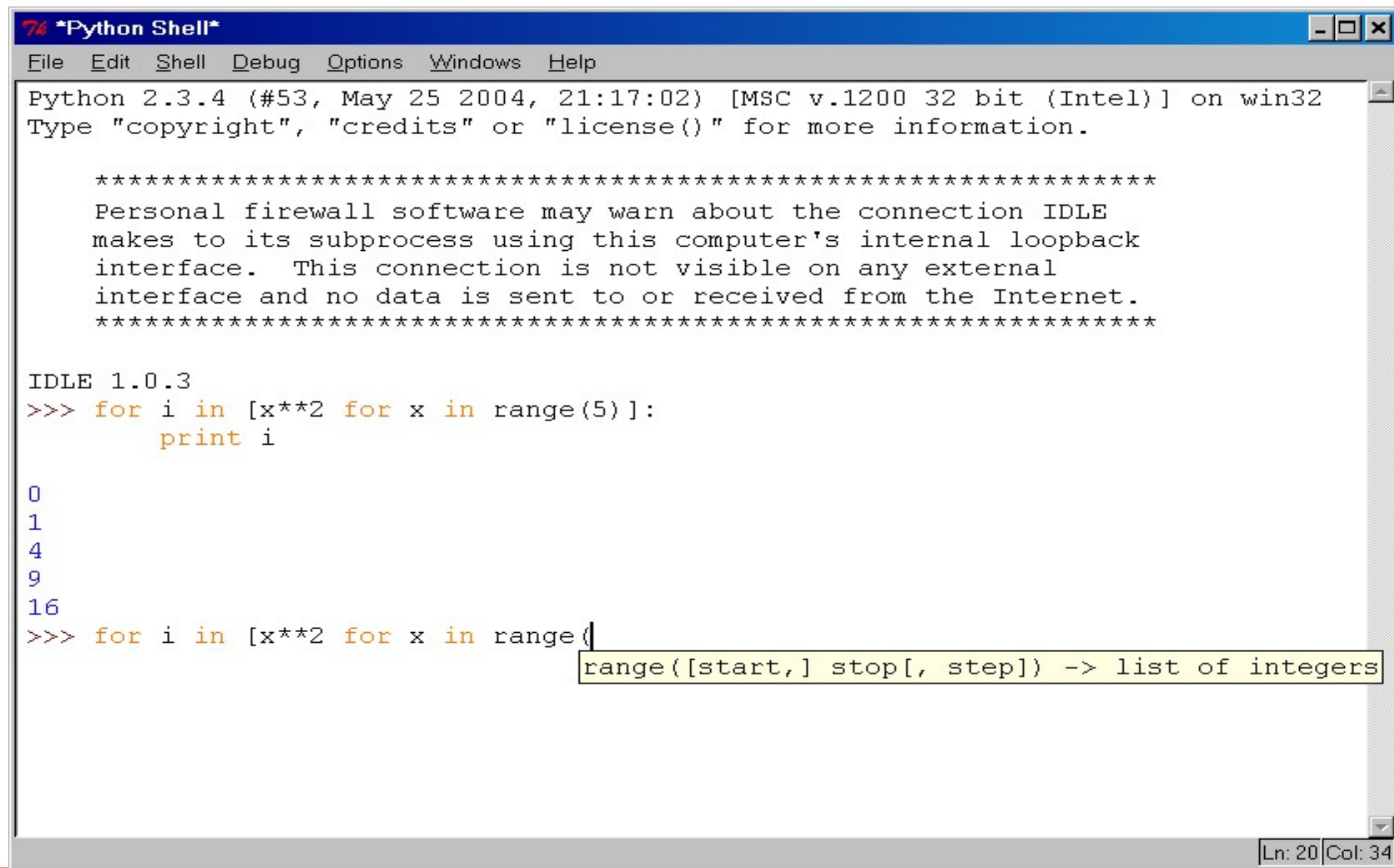
- A log entry showing 'hello world' from `index.js:1`.
- A command prompt `>` followed by the command `console.log('hello world')`.
- A log entry showing 'hello world' from `VM244:1`.
- A log entry showing `<` undefined.
- A new command prompt `>` with a cursor.

The bottom of the console shows the 'Console' tab and a close button.

Cửa sổ soạn thảo chương trình(.js file) của Kdevelop



Tiếp cận lại giữa interpreter và compiler: Python



```
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(
```

range([start,] stop[, step]) -> list of integers

Ln: 20 Col: 34

Các công cụ liên quan đến trình biên dịch

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

