



Công cụ cho Khoa học dữ liệu

Date: @March 21, 2024, Writer: @Khôi Nguyễn

Link Notion: [CCKHDL Notion](#)

Recall

▼ Ngôn ngữ lập trình cho KHDL

- Các ngôn ngữ chính ?
- Đặc trưng ?

▼ API

- Máy loại ?
- Chức năng ?

▼ Loại dữ liệu

- Máy loại ?

Notes

▼ Ngôn ngữ lập trình cho KHDL

- Có rất nhiều ngôn ngữ lập trình.
- Ngôn ngữ cho KHDL có 3 ngôn ngữ chính:

▼ Python

- Ngôn ngữ lập trình đa năng, dễ đọc, dễ học và dễ nhớ.
- Nhiều thư viện hỗ trợ

▼ Khoa học dữ liệu

- **Pandas** - Thao tác với dữ liệu có cấu trúc và dữ liệu thời gian.
- **Numpy** - Xử lý các mảng và ma trận, đa chiều.
- **Matplotlib** - Trực quan hoá dữ liệu.
- ...

▼ Trí tuệ nhân tạo

- **Scikit-learn** - Triển khai thuật toán học máy.
- **Tensorflow** - Xây dựng và huấn luyện mô hình học sâu.

- Kể tên và ví dụ

- **Keras** - Tạo và triển khai mô hình học sâu một cách nhanh chóng.
- **PyTorch (Torch)** - Xây dựng và huấn luyện mô hình học sâu.
- ...

▼ R

- Ngôn ngữ lập trình dành cho tính toán và đồ họa thống kê.
- Có chức năng tích hợp cho máy học và trực quan hóa dữ liệu.

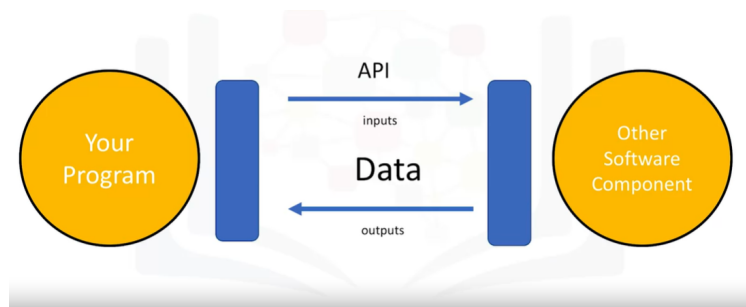
▼ SQL

- Ngôn ngữ truy vấn để thao tác với cơ sở dữ liệu.
- Cần thiết cho việc lưu trữ và phân tích dữ liệu có cấu trúc.
- Truy cập dữ liệu hiệu quả.
- VD: Microsoft SQL Server, My SQL, IBM DB2, PostgreSQL, Apache OpenOffice, SQLite, Oracle Database, MariaDB, ...

▼ API

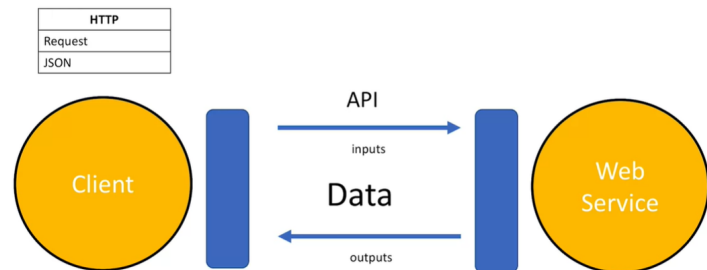
▼ API Interface

- Cho phép các phần mềm giao tiếp với nhau.



▼ Rest API

- Cho phép các phần mềm giao tiếp thông qua Internet.



▼ Loại dữ liệu

▼ Dữ liệu có cấu trúc

- Loại dữ liệu được tổ chức, phân biệt rõ ràng.

| Bảng Excel, Bảng trong cơ sở dữ liệu, ...

▼ Dữ liệu phi cấu trúc

- Loại dữ liệu không thể tổ chức hoặc phân biệt theo một quy tắc nhất định.

| Văn bản, hình ảnh, video, âm thanh, ...

▼ Dữ liệu bán cấu trúc

- Loại dữ liệu không tuân theo một cấu trúc cụ thể nhưng vẫn chứa một số cấu trúc mô tả, cho phép tổ chức và lưu trữ một cách linh hoạt.

| JSON, XML, ...

▼ — SUMMARY —

- Ngôn ngữ chính cho KHDL (Python, R, SQL).
- Các API để tương tác giữa các phần mềm với nhau.
- Các loại dữ liệu (có/phi/bán cấu trúc).