

ĐẠI HỌC BÁCH KHOA THÀNH PHỐ HỒ CHÍ MINH
KHOA KHOA HỌC VÀ KỸ THUẬT MÁY TÍNH



BÁO CÁO THỰC TẬP TỐT NGHIỆP

**Thư viện Web hỗ trợ tương tác dữ
liệu đa phương tiện**

Giáo viên hướng dẫn:

TS. Lương Thế Nhân

Sinh viên thực hiện:

Phạm Thành Công - 51200399

Đỗ Đặng Thanh Huy - 51201337

Thành phố Hồ Chí Minh, 05/2016

Mục lục

1	Giới thiệu đề tài	3
1.1	Bối cảnh	3
1.2	Mục tiêu	4
2	Các công trình liên quan	4
2.1	Công cụ hỗ trợ trực quan dữ liệu phân tán Gapminder	4
2.1.1	Giới thiệu	4
2.1.2	Mục tiêu	5
2.1.3	Phương pháp	6
2.1.4	Đánh giá	6
2.2	Global Health Atlas - WHO	7
2.2.1	Giới thiệu	7
2.2.2	Mục tiêu	8
2.2.3	Phương pháp	8
2.2.4	Đánh giá	9
2.3	Thư viện đồ thị C3.js	10
2.3.1	Giới thiệu	10
2.3.2	Mục tiêu	10
2.3.3	Phương pháp	12
2.3.4	Đánh giá	12
2.4	Ứng dụng tạo bản đồ online Mapme	13
2.4.1	Giới thiệu	13
2.4.2	Mục tiêu	13
2.4.3	Phương pháp	14
2.4.4	Đánh giá	16
3	Kiến thức nền	16
3.1	JavaScript	16
3.1.1	Giới thiệu	16

3.1.2	Mục tiêu	16
3.1.3	Tính năng	17
3.1.4	Ứng dụng	17
3.2	jQuery	17
3.2.1	Giới thiệu	17
3.2.2	Mục tiêu	19
3.2.3	Tính năng	19
3.2.4	Ứng dụng	19
3.3	d3.js	19
3.3.1	Giới thiệu	19
3.3.2	Mục tiêu	20
3.3.3	Ứng dụng	20
3.4	OpenLayers	21
3.4.1	Giới thiệu	21
3.4.2	Mục tiêu	21
3.4.3	Đặc điểm	21
3.4.4	Ứng dụng	22
4	Giải pháp đề xuất	23
4.1	Mô tả bài toán	23
4.2	Kiến trúc thư viện dự kiến xây dựng	23
4.3	Mô hình dữ liệu dự kiến xây dựng	25
5	Phương pháp đánh giá	26
6	Kế hoạch thực hiện	27
7	Kết luận	28
8	Tài liệu tham khảo	28

1 Giới thiệu đề tài

1.1 Bối cảnh

Trực quan dữ liệu, được hiểu là cách dùng hình ảnh để biểu diễn thông tin, ngày nay đã trở nên phổ dụng bởi những lợi ích nó mang lại.

Trực quan hóa khai thác hệ thống thị giác con người để cung cấp một cách trực quan, nhanh chóng, độc lập với ngôn ngữ để xem và hiển thị dữ liệu. Đây là công cụ đặc lực để phân tích và tìm hiểu thông tin. Cho đến nay, hệ thống thị giác con người là thứ đắt nhất, trực tiếp nhất, đem lại lượng thông tin lớn nhất lưu trong bộ nhớ con người. Lượng bộ nhớ tiêu tốn để xử lý thông tin thị giác vượt xa hơn so với những giác quan khác của con người. Một số nghiên cứu ước tính rằng hệ thống thị giác của con người có thể xử lý 9 megabit thông tin trên 1 giây, tương đương với gần 1 triệu kí tự trên 1 giây (2013)^[2].

Mặc cho tiềm năng của nó, Trực quan dữ liệu vẫn chưa được đánh giá cao bởi sự thiếu hiểu biết về nó. Nhiều xu hướng ngày nay về Trực quan dữ liệu thực sự đã gây ra những tác dụng phụ mang tính tiêu cực, đó là sự nhầm lẫn thay vì thấu hiểu thực sự về Trực quan dữ liệu. Ngày nay trong lĩnh vực kinh doanh hiệu quả (Business Intelligence) không có gì có thể mang chúng ta lại gần hơn với những hứa hẹn về sự tương tác thông minh hơn là Trực quan dữ liệu. Nhưng điều này chỉ xảy ra khi chúng ta thực sự hiểu và dùng nó một cách đúng đắn. Để đạt được điều đó, chúng ta phải thực sự hành động và vứt bỏ những quan niệm chưa đúng về Trực quan dữ liệu.

Trực quan dữ liệu ngày càng đóng vai trò quan trọng trong mọi lĩnh vực đời sống. Như sử dụng trong nghiên cứu, trong những công việc liên quan tới xử lý dữ liệu, giao dịch bởi người dùng phổ thông, và dùng với tỷ lệ ngày càng tăng trong lực lượng lao động trí óc, đặc biệt đối với các nhà phân tích. Đó là những tin mang tính khởi sắc. Bên cạnh đó, vẫn còn những mặt hạn chế, trong giới doanh nghiệp, Trực quan dữ liệu vẫn còn bị bỏ ngỏ, hiểu nhầm, sử dụng chưa hiệu quả, và thường bị làm sai lệch đi bởi các nhà cung cấp, sản xuất và bán phần mềm trực quan.

1.2 Mục tiêu

Thực quan dữ liệu đã trở thành một lĩnh vực nghiên cứu trong những năm gần đây. Nhiều trường đại học đã mở các Khoa chuyên nghiên cứu về Thực quan và có một số chương trình ưu việt phục vụ nhu cầu nghiên cứu và chế tạo các nguyên mẫu của sinh viên. Cộng đồng nghiên cứu này bao gồm các cá nhân đến từ nhiều lĩnh vực không chỉ riêng Khoa học Máy tính, như tâm thần học và thậm chí các doanh nghiệp, nơi tạo môi trường lý tưởng để phát triển lý thuyết Thực quan một cách thực tế và mang tính cách mạng nhất.

Chúng ta đang bắt đầu nhận thấy các sản phẩm Thực quan dữ liệu đang thực sự tiện ích như thế nào. Khi một công ty có thể phân tích big data, họ sẽ được nhiều lợi ích. Trong một cuộc khảo sát^[1], 63% cho rằng họ tin tưởng việc hiểu biết và khai thác big data 1 cách hiệu quả có thể tạo ra lợi thế cạnh tranh cho tổ chức. Phân tích big data có thể giúp họ cải thiện việc ra quyết định, tạo ra 1 cái nhìn toàn diện về khách hàng của họ, tăng tính bảo mật và kiểm soát, phân tích hoạt động và gia tăng kho dữ liệu. Thực quan hóa có vai trò quan trọng trong việc sử dụng big data để có 1 cái nhìn đầy đủ về khách hàng của họ.

Bởi vì những lợi ích mà Thực quan hoá mang lại, dựa trên những công trình nghiên cứu liên quan, mục tiêu của đề tài là xây dựng một thư viện hỗ trợ người dùng có thể dễ dàng tiếp cận và xây dựng một ứng dụng để thực quan dữ liệu. Đồng thời, so sánh đánh giá các phương pháp đã có để tìm ra mô hình thiết kế phù hợp.

2 Các công trình liên quan

2.1 Công cụ hỗ trợ thực quan dữ liệu phân tán Gapminder

2.1.1 Giới thiệu

Gapminder^[3] được thành lập tại Stockholm bởi Ola Rosling, Anna Rosling Rönnlund và Hans Rosling vào ngày 25 tháng 2 năm 2005.

Gapminder là một tổ chức phi lợi nhuận nhằm thúc đẩy sự phát triển bền vững toàn cầu, và là một thành tựu đánh dấu mục tiêu phát triển mang tầm thế kỷ của tổ chức Liên Hiệp Quốc (United Nations). Với mục tiêu đề ra nhằm gia tăng khả năng tìm hiểu, phân tích và đánh giá các thống kê, dữ liệu liên quan đến các vấn đề xã hội, tài chính và phát triển môi

trường ở địa phương, quốc gia, và tầm thế giới.

Gapminder hoạt động dựa trên nền tảng được xác định trước, xây dựng các dịch vụ bởi ban quản trị, đôi khi hợp tác với các dự án tại các trường đại học, các tổ chức quốc tế, cơ quan Nhà nước cũng như các tổ chức phi chính phủ.

2.1.2 Mục tiêu

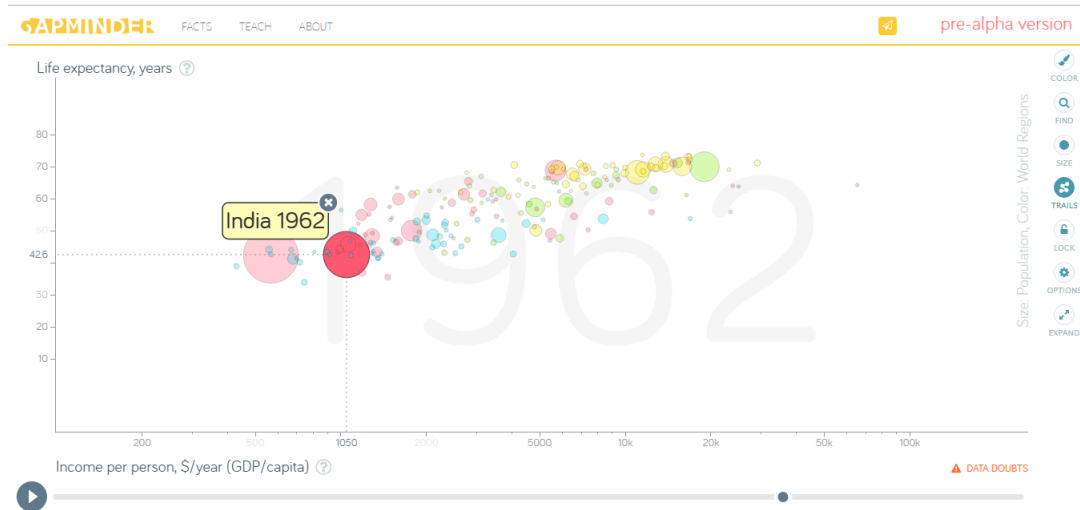
Các hoạt động ban đầu là theo đuổi sự phát triển của ứng dụng Trendalyzer (tháng 3 năm 2006). Trendalyzer là một phần mềm biểu diễn chuỗi thời gian thống kê bằng cách chuyển đổi những con số khô khan, nhàm chán thành các đồ thị chuyển động thú vị, và đặc biệt là khả năng tương tác. Trendalyzer còn được biết đến dưới tên gọi Gapminder World, xây dựng trên nền tảng web biểu diễn dữ liệu thống kê theo dòng thời gian cho tất cả các quốc gia trên thế giới.

Gapminder World^[4] cung cấp các video, bài thuyết trình bằng Flash và các đồ thị dưới dạng PDF thể hiện xu hướng phát triển mang tính toàn cầu.

Kể từ năm 2006 Gapminder cung cấp một lượng lớn các video thể hiện thực tế dưới góc nhìn toàn cầu.

Từ năm 2008 Gapminder còn cung cấp các biểu đồ con (sub-graph)^[5] thể hiện xu hướng, sự khác biệt giữa các tiểu bang của Mỹ, các tỉnh thành của Trung Quốc, hay các đồ thị đặc biệt so sánh số liệu thống kê từ các thành phố ở Trung Quốc, Mỹ, Ấn Độ với các quốc gia khác trên thế giới. Từ Gapminder World, chỉ bằng một cú click chuột, các dữ liệu, thống kê liên quan nằm sau các đồ thị được thể hiện trên các biểu đồ con.

Như ở hình 1, chúng ta có thể thấy với 1 lượng dữ liệu lớn (ở đây là độ tuổi kỳ vọng ở các quốc gia trên thế giới), có thể được thể hiện trực quan bởi đồ thị bong bóng (bubble chart). Qua kích thước các bong bóng, có thể thấy sự tương quan khác biệt giữa các quốc gia như thế nào. Hay khi trở vào 1 bong bóng, chúng ta thấy được số liệu chi tiết ở quốc gia đó, như trong hình là độ tuổi kỳ vọng ở Ấn Độ, vào năm 1962 là 42.6. Ngoài ra ở phía dưới đồ thị, có 1 thanh thể hiện dòng thời gian (timeline), khi thanh thời gian trượt đi, kích thước các quả bóng thay đổi theo như số liệu tương ứng ở các năm của các quốc gia.



Hình 1: Đồ thị trực quan dữ liệu Gapminder World

2.1.3 Phương pháp

Gapminder sử dụng D3.js, một thư viện JavaScript phổ biến trong việc thao tác với các thành phần của một trang web (Document Object Model) dựa trên dữ liệu.

D3 cho phép thể hiện dữ liệu thông qua HTML, SVG và CSS. D3 còn là một tiêu chuẩn web cung cấp hầu hết các khả năng mà một trình duyệt hiện đại mang lại, mà không cần phải tự viết lại hay dựa trên một nền tảng khuôn khổ. Kết hợp các thành phần trực quan mạnh mẽ và hướng tiếp cận dữ liệu để thao tác trên DOM.

2.1.4 Đánh giá

Với mục tiêu luôn luôn cập nhật các công cụ cũng như dữ liệu, nhằm đảm bảo tính xác thực, và đúng đắn của ứng dụng, Gapminder World đảm bảo nguồn dữ liệu chính xác nhất và cập nhật nhất. Dữ liệu đề tài sẽ sử dụng từ Gapminder để đảm bảo nguồn thông tin luôn xác thực và thể hiện đúng đắn nhất các xu hướng phát triển toàn cầu.

Ngoài ra, cách sử dụng các biểu đồ con để hiển thị dữ liệu liên quan của một đối tượng được tương tác giúp người dùng dễ dàng nắm bắt được chính xác các thông tin muốn truyền đạt, tăng tính thân thiện cũng như trải nghiệm người dùng. Áp dụng tính hiệu quả trên, đề tài đề xuất sử dụng cách thể hiện trải nghiệm người dùng (UI/UX) kế thừa các đặc tính của

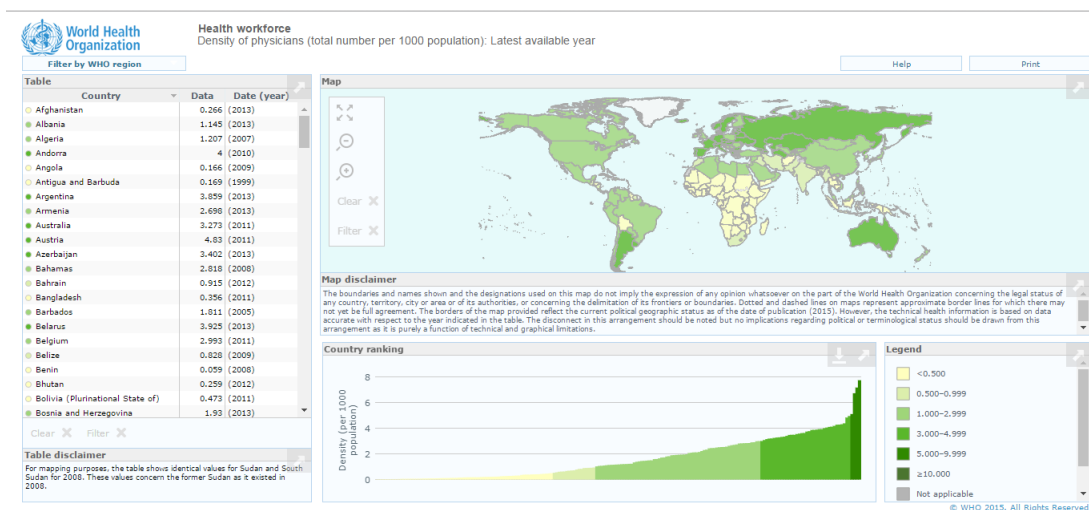
Gapminder World.

Về thư viện D3.js được sử dụng trong Gapminder World, đây là một thư viện mạnh mẽ tiếp cận dữ liệu dựa trên các thao tác với DOM. D3.js cho phép tái sử dụng các thành phần và trình nhúng (plugin) một cách đơn giản và hiệu quả. Đề tài đề xuất sử dụng D3.js làm thư viện chính để hiện thực thư viện tương tác giữa các đồ thị.

2.2 Global Health Atlas - WHO

2.2.1 Giới thiệu

Trong một môi trường điện tử, WHO's Communicable Disease Global Atlas được dùng cho việc phân tích và so sánh dữ liệu chuẩn, thống kê các bệnh truyền nhiễm ở các quốc gia, vùng lãnh thổ, và toàn thế giới. Việc phân tích và biểu diễn dữ liệu ngày càng được hỗ trợ tốt hơn thông qua các thông tin về dân số, điều kiện kinh tế xã hội, và các yếu tố môi trường. Với việc làm đó, biểu đồ Atlas này có thể cho biết được những yếu tố gây ra bệnh truyền nhiễm (2007)^[15].



Hình 2: Biểu đồ tương tác về mật độ bác sĩ của WHO

2.2.2 Mục tiêu

Ban đầu, với mục đích cung cấp những dữ liệu, báo cáo về các bệnh truyền nhiễm chính (sốt rét, HIV/AIDS,...), thì ngày nay Global Atlas của WHO cũng dùng để phân tích, biểu diễn, so sánh dữ liệu từ những vấn đề liên quan đến y tế khác như: mật độ bác sĩ, số lượng người chết do tai nạn giao thông, tuổi thọ trung bình,...

2.2.3 Phương pháp

Trực quan dữ liệu thành nhiều dạng biểu đồ khác nhau, các biểu đồ có thể tương tác với nhau giúp cho việc nhìn nhận, phân tích, đánh giá thông tin một cách chính xác, đầy đủ.

Ứng dụng^[16] bao gồm các thành phần:

- Bảng dữ liệu: được lấy từ database của WHO
- Map: dùng thư viện Google Maps, kết hợp với bảng dữ liệu để tạo ra biểu đồ
- Biểu đồ cột: dựa vào bảng dữ liệu để tạo ra biểu đồ
- Chú thích: phân vùng dữ liệu

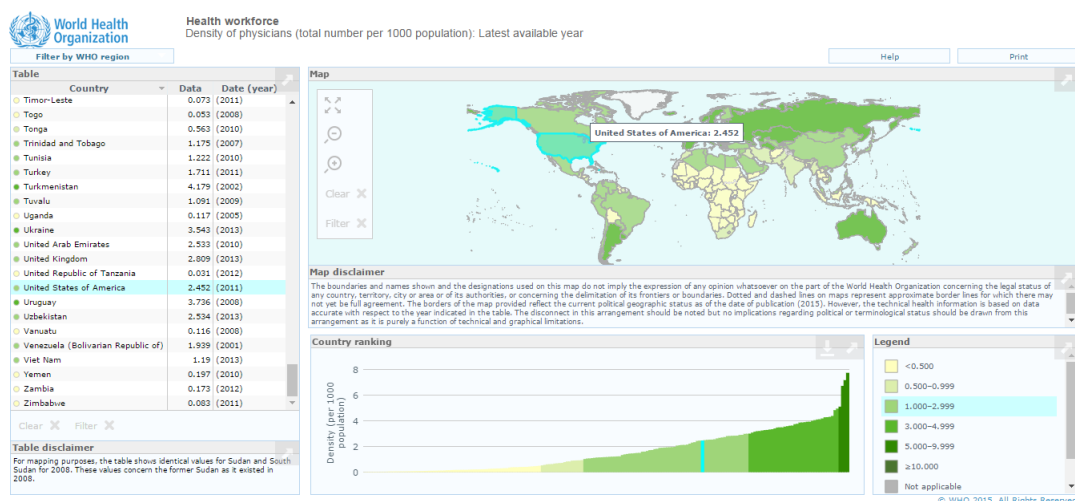
Tính năng: ứng dụng hỗ trợ tương tác giữa các biểu đồ, bảng dữ liệu:

- Khi click vào 1 hàng trong bảng dữ liệu, bên map sẽ phóng to đến vùng tương ứng, bên biểu đồ cột sẽ nổi bật cột tương ứng. Các tương tác sẽ được thể hiện tương tự khi click vào các biểu đồ.
- Khi hover lên 1 hàng trong bảng dữ liệu, bên các biểu đồ cũng sẽ được làm



Hình 3: Tương tác khi click vào một hàng trong bảng dữ liệu

nổi bật vùng tương ứng như khi click, nhưng với 1 màu sắc khác (bên map sẽ không phóng to đến vùng tương ứng).



Hình 4: Tương tác khi hover lên Map

2.2.4 Đánh giá

Dữ liệu được WHO thu thập trên toàn thế giới, cập nhật theo thời gian, đảm bảo tính chính xác của ứng dụng.

Bên cạnh đó, việc trực quan thành nhiều loại biểu đồ, thể hiện được nhiều cái nhìn khác nhau của dữ liệu giúp đánh giá vấn đề 1 cách đúng đắn, đầy đủ.

Một tính năng khác đó là việc tương tác giữa các đồ thị giúp người dùng có nhiều cái nhìn từ nguồn dữ liệu, hiểu được thông tin rút trích từ các đồ thị, tăng tính thân thiện, trải nghiệm của người dùng.

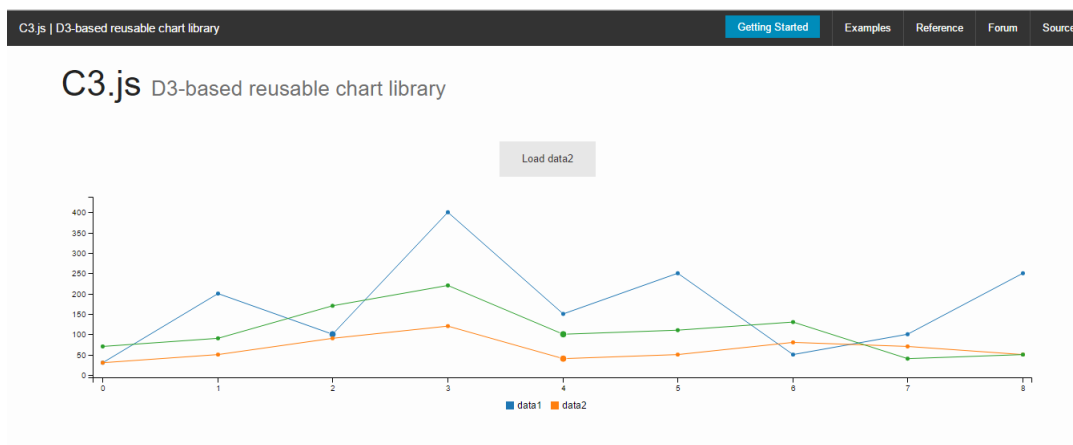
Do đó, đề tài đề xuất sử dụng nhiều đồ thị để biểu diễn cho nguồn dữ liệu đầu vào, và sự tương tác giữa các đồ thị.

2.3 Thư viện đồ thị C3.js

2.3.1 Giới thiệu

C3.js là một thư viện mã nguồn mở^[6] dựa trên thư viện D3.js , nhằm cho phép người dùng có cái nhìn sâu hơn về các ứng dụng web tích hợp đồ thị.

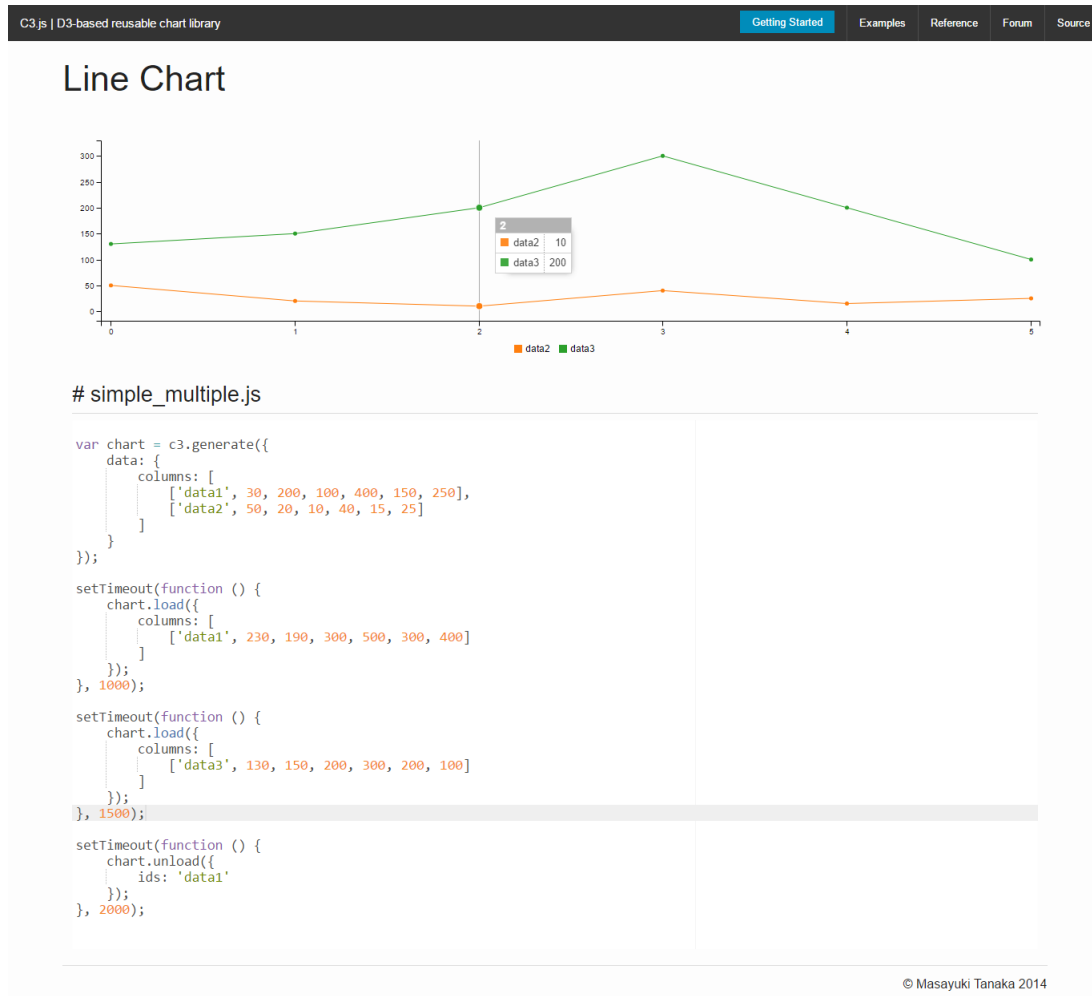
Toàn bộ mã nguồn của thư viện nằm ở: <https://github.com/masayuki0812/c3>



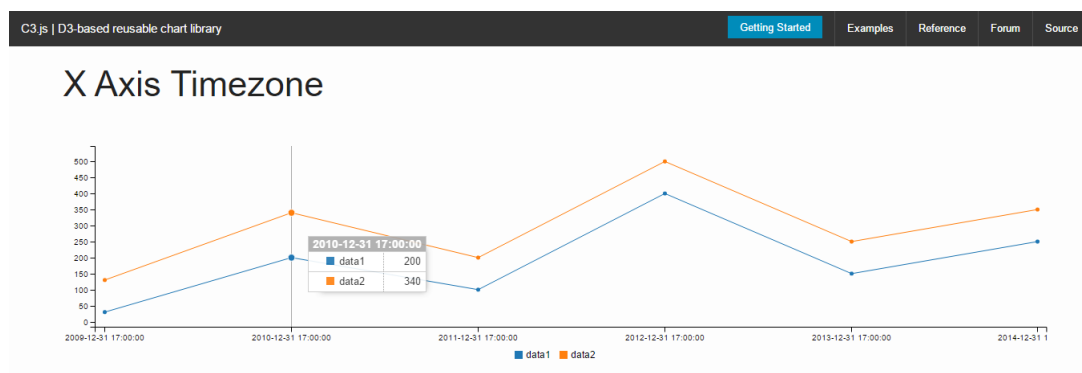
Hình 5: Thư viện đồ thị C3js

2.3.2 Mục tiêu

- Sự thoải mái (Comfortable): C3 giúp người dùng có thể dễ dàng tạo một đồ thị dựa trên D3 bằng cách gộp chung các đoạn code cần thiết để cấu trúc một đồ thị hoàn chỉnh. Qua đó, người dùng không cần phải viết các đoạn code D3 rườm rà như trước nữa. [Hình 6] Thông qua một số câu lệnh định nghĩa dữ liệu, C3 khởi tạo một đồ thị cho phép tương tác một cách trực quan thông qua hover, click,... các thành phần của đồ thị
- Khả năng tùy chỉnh (Customizable): C3 cung cấp các class tương ứng cho mỗi đối tượng trong đồ thị khi khởi tạo. Qua đó, người dùng có thể định nghĩa các style theo ý muốn theo class cũng như mở rộng cấu trúc các thành phần của đồ thị (trục dữ liệu, hình dạng hiển thị, kiểu đồ thị,..) một cách trực tiếp. [Hình 7] Trục hoành của đồ thị



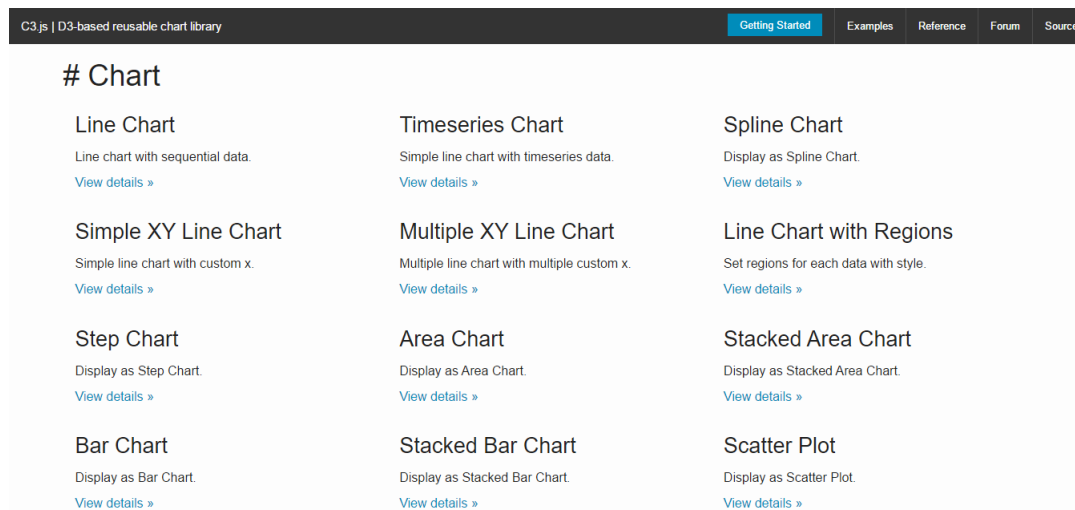
Hình 6: Thư viện đồ thị C3js



Hình 7: Thư viện đồ thị C3js

được tùy chỉnh theo thời gian hiện tại với định dạng hoàn toàn tùy thuộc người sử dụng.

- Khả năng kiểm soát (Controllable): C3 cung cấp đa dạng các APIs và các hàm callback để truy xuất trạng thái của đồ thị. Bằng cách đó, người dùng có thể cập nhật lại đồ thị thậm chí sau khi đồ thị đã được thể hiện. [Hình 8] Thể hiện một số thành phần được hỗ trợ bởi thư viện C3js.



Hình 8: Thư viện đồ thị C3js

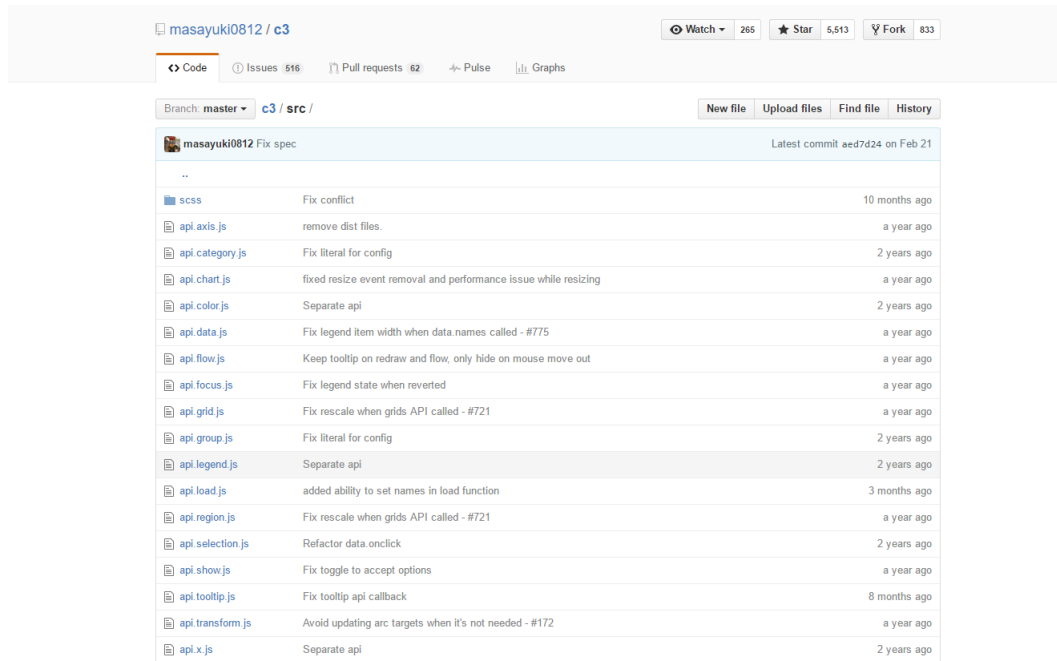
2.3.3 Phương pháp

Như đã trình bày ở trên, C3js sử dụng thư viện D3js nhằm tạo các đồ thị có khả năng tái sử dụng, cũng như giúp người dùng đào sâu hơn vào khái niệm trực quan dữ liệu thông qua đồ thị trên nền ứng dụng web.

Ngoài ra dựa theo cấu trúc mã nguồn mà C3js chia sẻ, có thể thấy cấu trúc tổ chức của C3js theo hướng hướng đối tượng, được tổ chức theo các class (trục, dữ liệu, đồ thị,..) và các phương thức, thuộc tính đi kèm class.

2.3.4 Đánh giá

Cách tổ chức và hiện thực theo hướng hướng đối tượng như trên giúp tiết kiệm số dòng mã, tận dụng khả năng tái sử dụng, đồng thời giúp việc mở rộng về sau rất hiệu quả.



Hình 9: Thư viện đồ thị C3js

Đề tài đề xuất sử dụng cách hiện thực API theo hướng như thư viện C3js.

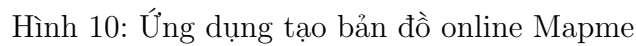
2.4 Ứng dụng tạo bản đồ online Mapme

2.4.1 Giới thiệu

Mapme^[7] là một ứng dụng cho phép người dùng tạo bản đồ, dữ liệu được xử lý trực tiếp từ các file Excel đầu vào. Bản đồ tạo ra có các marker tương ứng với dữ liệu, có khả năng tương tác trực tiếp thông qua các sự kiện định nghĩa trước. Ngoài ra, bản đồ được tạo ra còn có thể được nhúng vào các trang web khác thông qua một đoạn mã nhúng. Đặc biệt toàn bộ tương tác từ những bước khởi tạo bản đồ đầu tiên, đến lúc xuất thành quả, đều không yêu cầu người dùng phải viết một dòng mã nào.

2.4.2 Mục tiêu

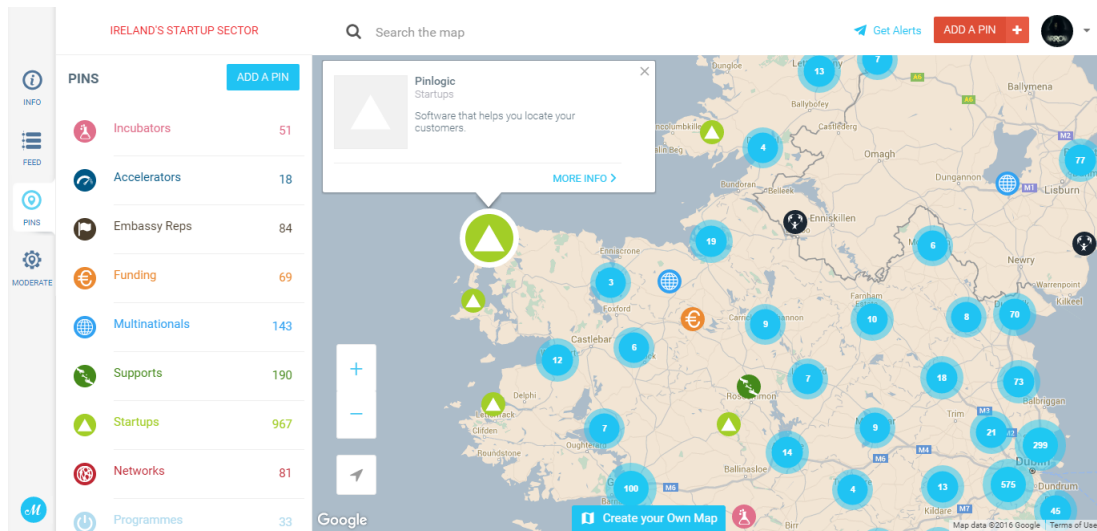
Đầu tiên nói đến các đối tượng mà Mapme hướng tới, đó là các tổ chức, chính phủ, các doanh nghiệp hay các tổ chức cộng đồng phi lợi nhuận.



2.4.3 Phương pháp

Như hình 11 có thể thấy, các thông tin về khởi nghiệp ở quốc gia Ireland được trực quan thông qua các biểu tượng (marker). Cùng với đó là các số liệu tương ứng được xử lý trực tiếp từ các file Excel đầu vào. Ngoài ra, người dùng còn có thể tương tác với các marker này để thể hiện thông tin chi tiết ở các điểm chỉ định.

Đặc biệt, Mapme còn hỗ trợ sinh mã nhúng [Hình 12], dưới dạng iframe, giúp người dùng nhúng trực tiếp bản đồ mình vừa tạo vào các website tùy mục đích.




Hình 11: Ứng dụng tạo bản đồ online Mapme

Embed sample-data-map

Copy/Paste the embed code in your website or blog.
[Learn more...](#)

```
<iframe src='https://mapme.com/sample-data-map' width='100%' height='500' marginwidth=0
marginheight=0 hspace=0 vspace=0 frameborder=0
scrolling='no' border=0></iframe>
```



Copy

Close

Hình 12: Ứng dụng tạo bản đồ online Mapme

2.4.4 Đánh giá

Đầu tiên là về cách sử dụng OpenStreetMap làm dịch vụ cung cấp bản đồ, đảm bảo ứng dụng luôn chính xác về các dữ liệu địa lý so với thực tế. Nhóm đề tài đề xuất sử dụng OpenStreetMap làm dịch vụ cung cấp bản đồ cho API mà nhóm đề tài thực hiện.

Ngoài ra, ứng dụng cho phép người dùng tương tác trực tiếp với các marker trên bản đồ, hiển thị các thông tin liên quan đến các địa điểm đó. Thay vì phải đọc các dữ liệu thô theo vùng miền, theo quốc gia, vùng lãnh thổ như trước, cách tương tác này giúp người dùng dễ dàng cập nhật thông tin cũng như có cái nhìn tổng quan, khả năng so sánh giữa các địa điểm khác nhau mà không phải chăm chú vào các con số nhàm chán như trước.

3 Kiến thức nền

Qua việc tìm hiểu cũng như những phần trình bày về các công trình liên quan đến đề tài, nhóm đề tài nhận thấy cần phải nắm được những kiến thức cơ bản sau để hỗ trợ tốt hơn cho việc hiện thực API về trực quan dữ liệu:

3.1 JavaScript

3.1.1 Giới thiệu

JavaScript^[19], theo phiên bản hiện hành mới nhất ECMAScript 6^[7], là một ngôn ngữ lập trình kịch bản (scripting language), đa nền tảng (cross-platform), nhỏ và nhẹ (lightweight). Ngôn ngữ này được dùng rộng rãi cho các trang web, nhưng cũng được dùng để tạo khả năng viết script sử dụng các đối tượng nằm sẵn trong các ứng dụng.

3.1.2 Mục tiêu

Mục tiêu của JavaScript là cung cấp cho các nhà phát triển Web một số khả năng và quyền điều khiển chức năng cho trang Web. Mã JavaScript có khả năng nhúng trong tài liệu HTML để điều khiển nội dung của trang Web và kiểm tra sự hợp lệ của dữ liệu mà người dùng nhập vào. Khi một trang hiển thị trong trình duyệt, các câu lệnh được trình duyệt thông dịch và thực thi.

3.1.3 Tính năng

JavaScript nâng cao sự linh động và khả năng tương tác cho website bằng cách sử dụng các hiệu ứng của nó như thực hiện các phép tính, kiểm tra form, viết các trò chơi, bổ sung các hiệu ứng đặc biệt, tùy biến các chọn lựa đồ họa, tạo ra các mật khẩu bảo mật và hơn thế nữa.

Cụ thể, JavaScript có thể được dùng để:

- Tương tác với người dùng: Chúng ta có thể viết mã để đáp lại các sự kiện. Các sự kiện này sẽ có thể phát sinh bởi người dùng: nhấp chuột hay được phát sinh từ hệ thống, định lại kích thước của trang ...
- Thay đổi nội dung động: Mã JavaScript có thể dùng để thay đổi nội dung và vị trí các phần tử một cách động trên một trang nhằm đáp lại sự tương tác với người dùng.
- Kiểm tra tính hợp lệ dữ liệu: Chúng ta có thể viết mã nhằm kiểm tra tính hợp lệ của dữ liệu do người dùng nhập vào trước khi nó được gửi lên Web server để xử lý.

3.1.4 Ứng dụng

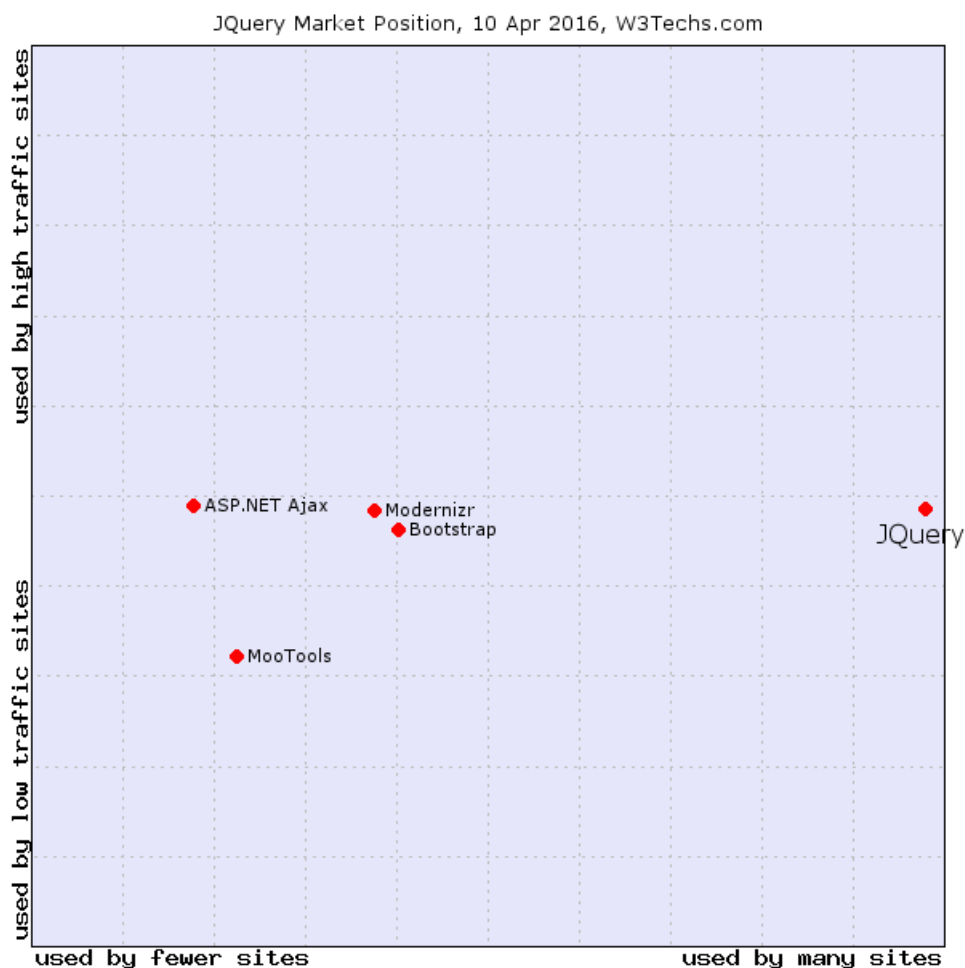
Dựa trên những đặc điểm và khả năng mạnh mẽ trong hỗ trợ lập trình API của JavaScript đã kể trên, nhóm đề xuất sử dụng JavaScript làm ngôn ngữ chính để hiện thực API. Ngoài ra, nhóm đề xuất sử dụng chuẩn mới nhất, ECMAScript 6 làm tiêu chuẩn cho các ràng buộc, quy tắc (coding conventions) khi hiện thực API.

3.2 jQuery

3.2.1 Giới thiệu

jQuery^{[12][13]} là một thư viện đa nền tảng (cross-platform) mã nguồn mở dựa trên ngôn ngữ JavaScript, được tạo bởi John Resig vào năm 2006. jQuery là một thư viện JavaScript phổ biến nhất được sử dụng hiện nay.

Theo thống kê (2016)^[11], cho tới hiện nay thì jQuery được sử dụng bởi 96% của tất cả các trang web có sử dụng thư viện JavaScript.



Hình 13: Vị trí của jQuery so với những thư viện JavaScript khác

3.2.2 Mục tiêu

Với phương châm là “write less, do more”, jQuery được thiết kế nhằm đơn giản hóa các nhiệm vụ khác nhau bằng cách viết ít mã hơn.

Thành phần quan trọng của jQuery là thư viện thao tác DOM (Document Object Model). DOM là 1 biểu diễn cấu trúc cây của tất cả các thành phần của 1 trang web và jQuery đơn giản hóa các cú pháp cho việc tìm kiếm, chọn lựa, và thao tác những phần tử DOM.

3.2.3 Tính năng

- Việc lựa chọn các phần tử DOM sử dụng công cụ Sizzle (selector mã nguồn mở)^[14]
- Thao tác với DOM dựa trên CSS Selector
- Sự kiện (event)
- Tạo chuyển động
- Hỗ trợ AJAX
- Có thể mở rộng thông qua các plug-in
- Được hỗ trợ bởi nhiều trình duyệt
- Cập nhật và hỗ trợ các công nghệ web mới nhất (HTML5 & CSS3)

3.2.4 Ứng dụng

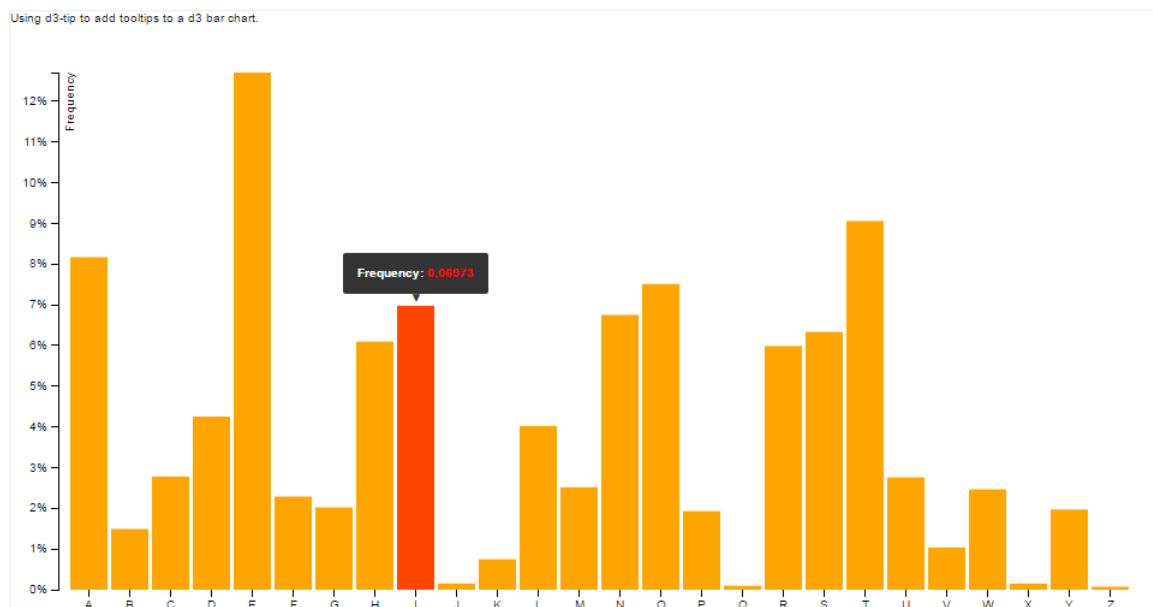
Dựa trên những đặc điểm cũng như tính năng mạnh mẽ của jQuery nói trên, đề tài sẽ đề xuất sử dụng jQuery như là 1 ngôn ngữ chính, tạo ra một API dễ dàng tìm hiểu, sử dụng với một số lượng mã (code) ít nhất để có thể tạo được ứng dụng của riêng mình.

3.3 d3.js

3.3.1 Giới thiệu

D3.js^[17]. được viết bởi Mike Bostock, là 1 thư viện JavaScript dùng để thao tác với document dựa trên dữ liệu (data). D3 giúp mang dữ liệu đến cuộc sống bằng cách sử dụng HTML,

SVG, và CSS. D3 trên chuẩn web cung cấp đầy đủ các tính năng của những trình duyệt hiện đại mà không buộc mình vào một khuôn khổ nhất định nào, kết hợp với các thành phần trực quan mạnh mẽ và một phương pháp hướng dữ liệu để thao tác DOM.



Hình 14: Một ví dụ về d3.js

3.3.2 Mục tiêu

D3.js giúp nhúng dữ liệu vào trong các phần tử DOM. Sau đó, có thể áp dụng CSS3, HTML, và/hoặc SVG lên dữ liệu này. Cuối cùng, D3.js có thể giúp tạo ra dữ liệu tương tác thông qua việc sử dụng những sự biến đổi (transformation), chuyển đổi (transition) hướng dữ liệu của D3.js

3.3.3 Ứng dụng

D3.js được viết trong JavaScript, do đó có thể sử dụng lại code và thêm những hàm chức năng mà người dùng muốn. Cách người dùng chọn để định dạng, thao tác, tương tác với dữ liệu là tùy thuộc ở họ.^[18]

D3.js rất thích hợp dùng trong đề tài, nó giúp xây dựng framework trực quan dữ liệu mà người dùng muốn, dễ dàng sử dụng, do đó giúp tạo ra API một cách dễ dàng hơn, API dễ

hiều, dễ sử dụng.

3.4 OpenLayers

3.4.1 Giới thiệu

OpenLayers^[20] được tạo ra bởi MetaCarta sau hội nghị O'Reilly Where 2.0 (29-30/06/2005) và chính thức phát hành thành mã nguồn mở trước hội nghị O'Reilly Where 2.0 (13-14/06/2006) bởi MetaCarta Labs. Hai dự án khác về công cụ xây dựng bản đồ phát hành bởi MetaCarta là FeatureServer và TileCache. Từ tháng 11/2007, OpenLayers trở thành dự án mã nguồn mở chính thức của Hội đồng Không gian địa lý.

Sơ qua về OpenLayers, đây là một thư viện JavaScript mã nguồn mở (được chứng nhận bởi BSD License) nhằm phục vụ mục đích biểu diễn thông tin trên bản đồ đối với các trình duyệt Web, không phụ thuộc phía server. Nó cung cấp API chuyên về xây dựng các ứng dụng địa lý trên nền Web tương tự như Google Maps^[21] và Bing Maps^[22], với một điểm khác biệt duy nhất, OpenLayers hoàn toàn miễn phí, được xây dựng và phát triển nên bởi cộng đồng phát triển phần mềm mã nguồn mở. Thư viện xây dựng trên Prototype JavaScript Framework^[10]. Hiện nay, OpenLayers đã ra tới phiên bản OpenLayers 3^[23]. OpenLayers 3 được viết lại nhằm dễ hiểu hơn, hướng tới các tính năng của HTML5 và CSS3. Thư viện tiếp tục hỗ trợ trình chiếu, các giao thức tiêu chuẩn và các chức năng chỉnh sửa. Phiên bản mới này tập trung vào cải thiện hiệu năng, xây dựng đơn giản hơn, các thành phần trở nên trực quan hơn, API được cải tiến và còn nhiều hơn nữa.

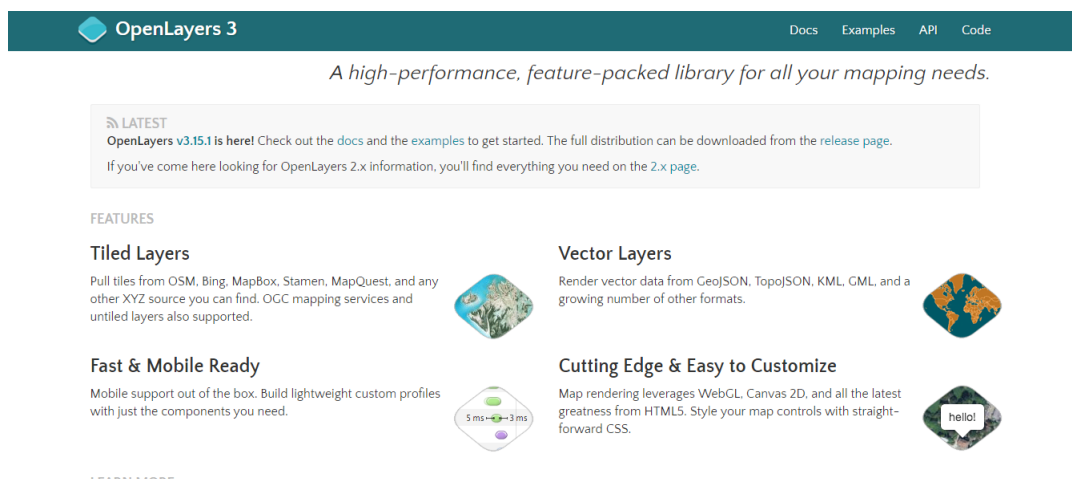
Mã nguồn của thư viện nằm ở địa chỉ sau: <https://github.com/openlayers/ol3/>

3.4.2 Mục tiêu

Như đã trình bày ở trên, OpenLayers hỗ trợ các lập trình viên hiển thị thông tin dưới dạng bản đồ, hay bất cứ nhu cầu nào liên quan về bản đồ một cách hiệu quả và đơn giản nhất.

3.4.3 Đặc điểm

OpenLayers có 4 đặc điểm chính:



Hình 15: Các đặc điểm chính của OpenLayers

- Thể hiện Lớp dưới dạng ô (Tiled Layers): Các lớp này được kéo về từ OpenStreetMap (OSM), Bing, MapBox, Stamen, MapQuest và bất cứ nguồn nào mà bạn có thể tìm thấy. Hỗ trợ đa dạng các Layer mà bạn có thể sử dụng trong ứng dụng của mình.
- Thể hiện Lớp dưới dạng Vector (Vector Layers): Sinh ra dữ liệu dạng vector từ GeoJSON, TopoJSON, KML, GML và một số định dạng khác.
- Tính đối ứng nhanh gọn (Fast & Mobile Ready): Hỗ trợ tùy chỉnh các thành phần một cách nhanh gọn và thuận tiện nhất. Cho phép xây dựng một bản tùy chỉnh gọn nhẹ gồm các thành phần mà lập trình viên cần.
- Mài dũa các góc cạnh và dễ dàng tùy chỉnh (Cutting Edge & Easy to Customize): Bản đồ sinh ra bởi công nghệ WebGL, Canvas 2D, và từ những tính năng ưu việt nhất của HTML5. Dễ dàng tùy chỉnh style bằng CSS.

3.4.4 Ứng dụng

Đề tài hướng tới xây dựng thư viện JavaScript cho phép tương tác giữa bản đồ và các đồ thị dựa trên dữ liệu đầu vào, cho nên OpenLayers, cụ thể OpenLayers 3 sẽ là công cụ đắc lực cho việc xây dựng thành phần bản đồ trong đề tài.

4 Giải pháp đề xuất

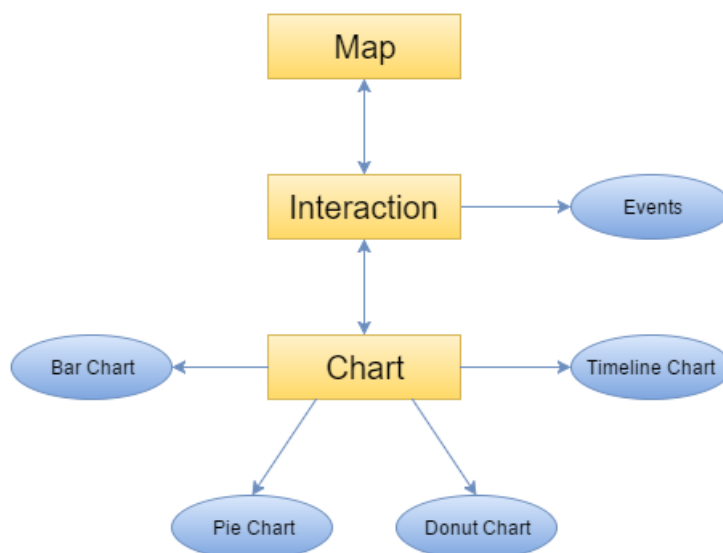
4.1 Mô tả bài toán

Qua các kiến thức thu được từ các công trình nghiên cứu liên quan đã tìm hiểu, nhóm đề xuất bài toán: "Xây dựng thư viện hỗ trợ phát triển các ứng dụng nền tảng Web trong tương tác dữ liệu đa phương tiện". Nhóm nhận thấy các công trình liên quan hiện tại chủ yếu hỗ trợ vào một thể hiện dữ liệu nhất định (hoặc dạng đồ thị, hoặc dạng bản đồ) mà chưa có công trình nào tập trung vào yếu tố tương tác giữa các mô hình này, dẫn đến sự tương quan trong việc biểu diễn các loại dữ liệu giảm tính hiệu quả cũng như chưa thể thoát hẳn khỏi các cách biểu diễn truyền thống. Hoặc là đã xây dựng các mô hình trực quan cho phép tương tác giữa các loại mô hình (điển hình như công trình Global Health Atlas - WHO đã trình bày ở trên) nhưng công nghệ đã cũ và không tự sinh thư viện hay mã nguồn hỗ trợ người dùng có thể tự xây dựng ứng dụng tương tự.

Do đó, nhóm nhận thấy bài toán đề xuất phù hợp để thực hiện và phát triển.

4.2 Kiến trúc thư viện dự kiến xây dựng

Qua bài toán đề xuất, nhóm nhận thấy cần xây dựng mô hình với 3 thành phần chính như hình 16:

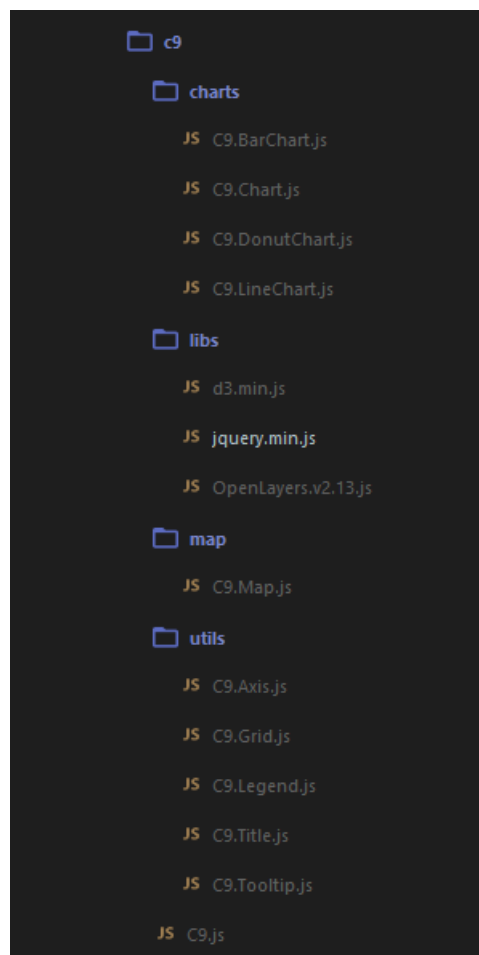


Hình 16. Mô hình dự kiến xây dựng

Với 3 thành phần (Component) chính là Map, Chart và Interaction, mỗi thành phần thể hiện rõ ràng chức năng tương ứng:

- Map: Map là một đối tượng biểu diễn dữ liệu dưới dạng bản đồ
- Chart: Chart biểu diễn dữ liệu đầu vào dưới dạng các dạng đồ thị định nghĩa trước (nhóm đề xuất một số dạng đồ thị phổ biến như đồ thị dạng thanh, dạng tròn, dạng mảnh và dạng dòng thời gian) phụ thuộc kiểu dữ liệu đầu vào cung cấp bởi người dùng.
- Interaction: Liên kết hai thành phần thể hiện dữ liệu trên là lớp Interaction, cung cấp các sự kiện hỗ trợ tương tác như Click, Hover trên từng đối tượng dữ liệu; Highlight các Component tương ứng theo các sự kiện được định nghĩa bởi người dùng.

Hình 17 là kiến trúc mã nguồn đề tài dự kiến hiện thực, gồm các file mã nguồn JavaScript được tổ chức tương ứng theo các thành phần đã định nghĩa ở trên:



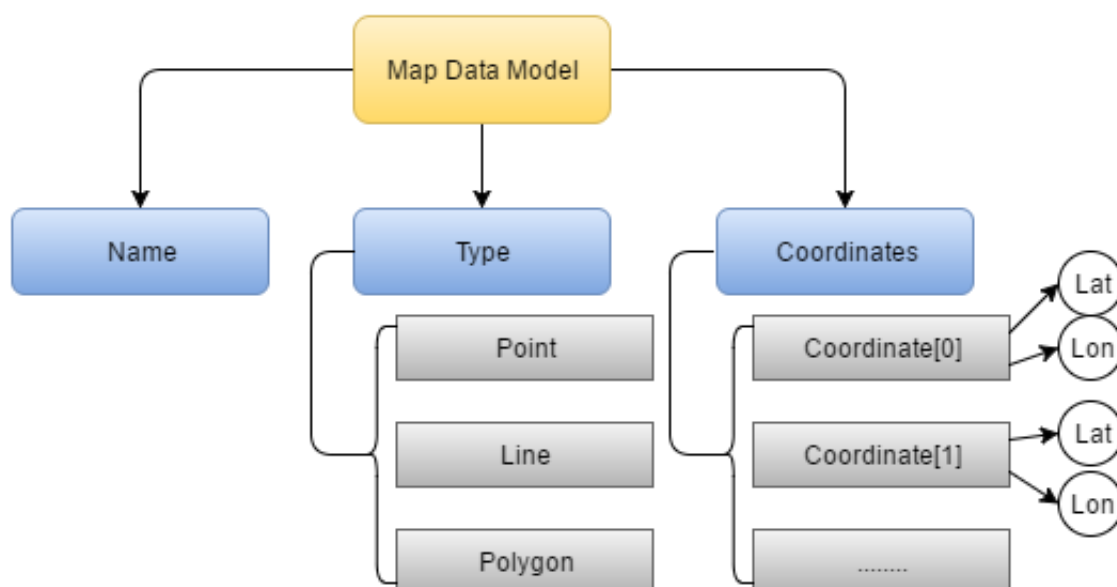
Hình 17. Kiến trúc mã nguồn dự kiến hiện thực

4.3 Mô hình dữ liệu dự kiến xây dựng

- Dữ liệu dạng bản đồ: Dữ liệu được tham khảo từ nguồn: <https://mapzen.com/data/borders>. Dữ liệu bản đồ được thể hiện dạng JSON, tổ chức theo các mức (administrative level) từ 1 tới 10, tương ứng từ biên giới quốc gia tới tỉnh, thành phố trực thuộc trung ương, quận, huyện, thị xã, thành phố thuộc tỉnh đến phường, xã, thị trấn.

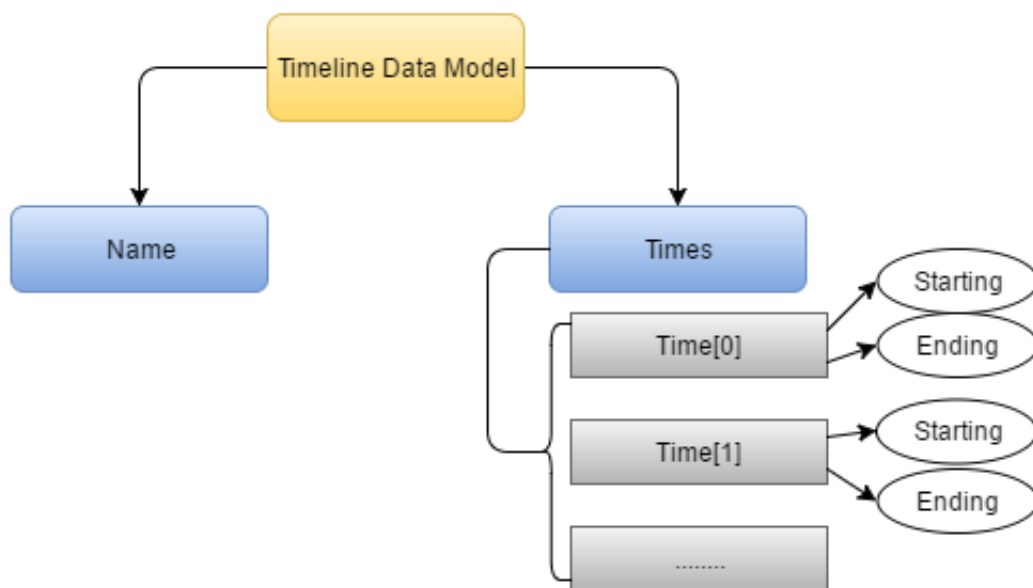
Đề tài dự kiến hỗ trợ bản đồ Việt Nam ở mức độ 4 (tỉnh, thành phố trực thuộc trung ương)

Mô hình chuẩn hóa dữ liệu bản đồ từ người dùng:



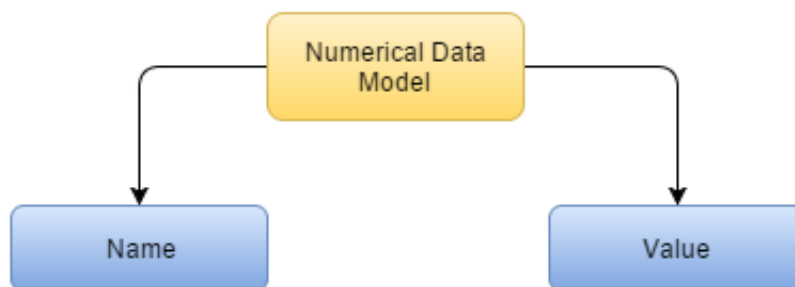
Hình 18. Mô hình dữ liệu bản đồ dự kiến xây dựng

- Dữ liệu dạng thời gian: Đối với dạng dữ liệu thời gian, đề tài đề xuất mô hình chuẩn hóa dữ liệu đầu vào từ người dùng như hình 19:



Hình 19. Mô hình dữ liệu thời gian dự kiến xây dựng

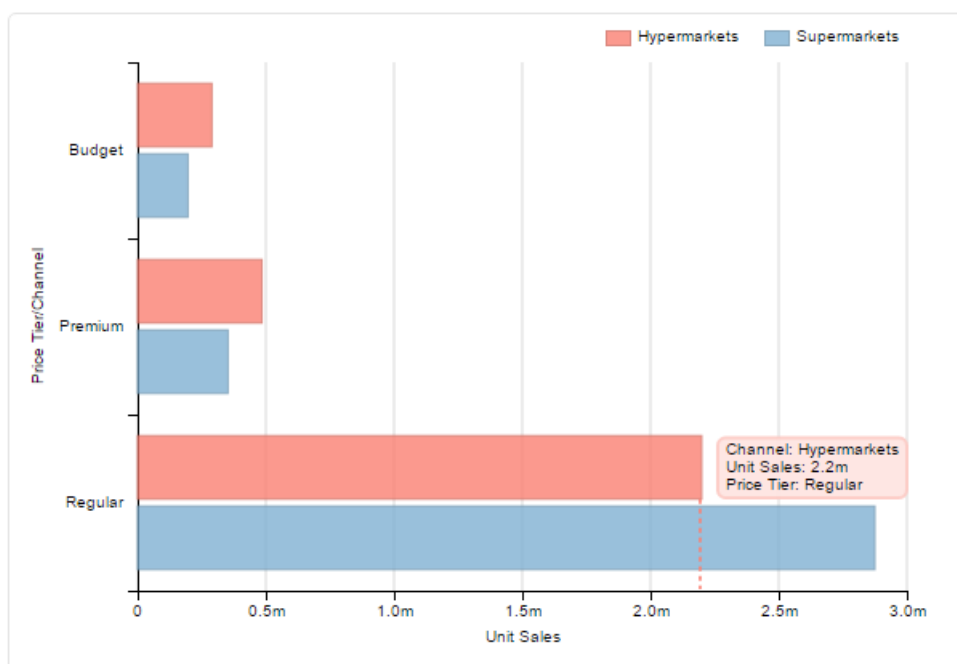
- Dữ liệu dạng số: Dữ liệu dạng số được thể hiện theo các cặp (name, value) như hình 20:



Hình 20. Mô hình dữ liệu dạng số dự kiến xây dựng

5 Phương pháp đánh giá

Bên cạnh việc thực hiện API, đề tài cũng thực hiện chương trình xây dựng một trang web mẫu sử dụng API làm ra để thể hiện tương tác các loại dữ liệu khác nhau.



Hình 21. Ví dụ về một output của API đã xây dựng

- Đầu vào là API đã được xây dựng mẫu
- Đầu ra là thể hiện của dữ liệu đồ thị sau khi sử dụng API.

6 Kế hoạch thực hiện

Trong giai đoạn thực tập tốt nghiệp, nhóm đã thực hiện được:

- Tìm hiểu về các công trình liên quan đến các thư viện web cho phép biểu diễn dữ liệu thành các dạng biểu đồ, bản đồ, và sự tương tác giữa chúng
- Đề xuất bài toán cụ thể nhóm sẽ thực hiện
- Tìm hiểu về các công cụ hỗ trợ cho việc thực hiện đề tài
- Hiện thực được API cho phần biểu đồ

Kế hoạch dự kiến thực hiện trong giai đoạn luận văn tốt nghiệp:

- Chuẩn hóa dữ liệu cho các dạng bản đồ, biểu đồ

- Hiện thực một thư viện cho phép người dùng có thể tự tạo ra một ứng dụng web tương tác dựa trên các loại dữ liệu khác nhau.
- Đánh giá và so sánh thư viện đã hiện thực được với những thư viện hiện có
- Tổng hợp và viết báo cáo

7 Kết luận

Năm 2006, Michael Palmer có viết “Data is the new oil”, điều đó có nghĩa rằng dữ liệu chỉ ở dạng thô, nó có giá trị, tuy nhiên nếu chưa qua sàng lọc thì nó không thể được sử dụng. Vì vậy dữ liệu phải được khai phá, phân tích để có thể có được giá trị thật sự.

Dữ liệu ngày nay rất phong phú, đa dạng. Để có thể phân tích được nó không phải là 1 việc dễ dàng. Có 1 câu thành ngữ: “A Picture is worth 1000 words”, nghĩa là 1 bức tranh có trị giá ngàn từ. Điều đó đã nói lên sức mạnh của việc phân tích dữ liệu thành dạng hình ảnh trực quan, sinh động, để con người có thể dễ dàng nắm bắt và hiểu được.

Do đó, với tầm quan trọng của việc trực quan hóa dữ liệu, nhóm đã quyết định chọn đề tài này. Nhóm cũng xin cảm ơn thầy Lương Thế Nhân đã giúp đỡ, tận tình hướng dẫn nhóm thực hiện đề tài. Nhóm xin hứa sẽ quyết tâm hoàn thành tốt đề tài này.

8 Tài liệu tham khảo

Tài liệu

- [1] Steve LaValle, Eric Lesser, Rebecca Shockley, Michael S. Hopkins and Nina Kruschwitz, MIT. Big Data, Analytics and the Path From Insights to Value, 2015. <http://search.proquest.com/openview/863e390a27e9a69d59ee5ffe2204f293/1?pq-origsite=gscholar>
- [2] T.Alan Keahey, IBM. Using visualization to understand big data, 2013. www.iotdataviz.com/meetup/1/YTW03349USEN.PDF

- [3] Gapminder. About Gapminder, 22, 5, 2016. <http://www.gapminder.org/about-gapminder/our-mission/>
- [4] Gapminder World?. How may I use Gapminder World?. http://www.gapminder.org/faq_frequently_asked_questions/#13
- [5] Gapminder Sub-graph. Examples about Gapminder sub-graph. https://www.gapminder.org/tools/bubbles#_state_marker_size
- [6] c3js. C3.js | D3-based reusable chart library, 22, 5, 2016. <http://c3js.org/>
- [7] Maps built with Mapme. Plot your Story - Combine maps, photos, videos & words. <http://mapme.com/gallery/>
- [8] OpenStreetMap. The project that creates and distributes free geographic data for the world. <https://www.openstreetmap.org/>
- [9] Prototype JavaScript Framework. A foundation for ambitious web applications, 2, 2005. https://en.wikipedia.org/wiki/Prototype_JavaScript_Framework
- [10] ECMAScript 6 — New Features: Overview & Comparison. Copyright (c) 2015-2016 Ralf S. Engelschall. <http://es6-features.org/>
- [11] W3Techs. Usage statistics and market share of JQuery for websites, 10, 4, 2016. <https://w3techs.com/technologies/details/js-jquery/all/all>
- [12] jQuery. jQuery, 10, 4, 2016. <https://jquery.com/>
- [13] Wikipedia. jQuery, 22, 5, 2016. <https://en.wikipedia.org/wiki/JQuery>
- [14] Wiki. Sizzle. <https://github.com/jquery/sizzle/wiki>
- [15] World Health Organization (WHO). Global Health Atlas, 10, 4, 2016. <http://apps.who.int/globalatlas/>
- [16] WHO. Health workforce: Density of physicians, 10, 4, 2016. http://gamapserver.who.int/gho/interactive_charts/health_workforce/PhysiciansDensity_Total/atlas.html

- [17] D3.js. D3.js - Data-Driven Documents, 10, 4, 2016. <https://d3js.org/>
- [18] Dashing D3.js. Why build Data Visualization with D3.js, 10, 4, 2016. <https://www.dashingd3js.com/why-build-with-d3js>
- [19] Mozilla Developer Network. Introduction - JavaScript, 10, 4, 2016. <https://developer.mozilla.org/en-US/docs/Web/JavaScript/Guide/Introduction>
- [20] Open Layers 3. A high-performance, feature-packed library for all your mapping needs. <http://openlayers.org/en/latest/doc/>
- [21] Google Maps. Web mapping service developed by Google, 10, 2004. https://en.wikipedia.org/wiki/Google_Maps
- [22] Bing Maps. Web mapping service provided as a part of Microsoft's Bing suite of search engines and powered by the Bing Maps for Enterprise framework, 12, 5, 2010. https://en.wikipedia.org/wiki/Bing_Maps
- [23] Open Layers 3. Release Builds <http://openlayers.org/download/>